ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

GLOBAL CONTINUATION FOR DISTANCE GEOMETRY PROBLEMS

Jorge J. Moré and Zhijun Wu

Mathematics and Computer Science Division

Preprint MCS-P505-0395

March 1995

Work supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38 and by the Argonne Director's Individual Investigator Program.

ABSTRACT

Distance geometry problems arise in the interpretation of NMR data and in the determination of protein structure. We formulate the distance geometry problem as a global minimization problem with special structure, and show that global smoothing techniques and a continuation approach for global optimization can be used to determine solutions of distance geometry problems with a nearly 100% probability of success.

GLOBAL CONTINUATION FOR DISTANCE GEOMETRY PROBLEMS

Jorge J. Moré and Zhijun Wu

1 Introduction

A molecule with m atoms can be described by specifying the positions $x_1, \ldots, x_m$ in $\mathbb{R}^3$ of all the atoms in the molecule. If we are given bond lengths $\delta_{i,j}$ between a subset $\mathcal{S}$ of the atom pairs, it is important to determine whether there is a molecule that satisfies these bond length constraints. We pose this problem in terms of finding $x_1, \ldots, x_m$ such that

$$\|x_i - x_j\| = \delta_{i,j}, \qquad (i,j) \in \mathcal{S}. \tag{1.1}$$

If there is no solution $x_1, \ldots, x_m$ to these constraints, then the bond length specification must be in error. This can happen, for example, if the triangle inequality

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

is violated for atoms $\{i, j, k\}$ with bond length constraints.

Distance geometry problems that arise in the interpretation of NMR data and in the determination of protein structure are usually associated with the more general problem of finding positions $x_1, \ldots, x_m$ in $\mathbb{R}^3$ such that

$$l_{i,j} \le \|x_i - x_j\| \le u_{i,j}, \qquad (i,j) \in \mathcal{S}, \tag{1.2}$$

where $l_{i,j}$ and $u_{i,j}$ are lower and upper bounds on the distance constraints, respectively. For surveys of work in this area, see Crippen and Havel [4], Havel [9], Kuntz, Thomason, and Oshiro [16], and Brünger and Nilges [1]. We do not consider the general problem (1.2) because the aim of this paper is to show that algorithms based on the continuation approach for global optimization can be used to determine solutions of (1.1) with a nearly 100% probability of success. The techniques of this paper can be extended to (1.2), but the theory is not as elegant.

Distance geometry problems are NP-hard. Crippen and Havel [4] proved this result when all the atoms are restricted to $\mathbb{R}^1$ by reducing the distance geometry problem to the set partition problem: Given positive integers $s_1, \ldots, s_m$, determine a partition of these integers in sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that

$$\sum_{i \in \mathcal{S}_1} s_i = \sum_{i \in \mathcal{S}_2} s_i.$$

Work supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38 and by the Argonne Director's Individual Investigator Program.

The proof is instructive. Given an instance of the set partition problem, consider a distance geometry problem in $\mathbb{R}^1$ with $m+1$ atoms, where

$$\delta_{i,i+1} = s_i, \quad 1 \le i \le m, \qquad \delta_{1,m+1} = 0.$$

If the distance geometry problem has a solution, then $\delta_{1,m+1} = 0$ implies that $x_{m+1} = x_1$, and thus

$$\sum_{i=1}^{m} (x_{i+1} - x_i) = x_{m+1} - x_1 = 0.$$

Since $|x_{i+1} - x_i| = s_i$, the sets $S_1 = \{i : x_{i+1} - x_i \ge 0\}$ and $S_2 = \{i : x_{i+1} - x_i < 0\}$ solve the set partition problem. For a general discussion of the complexity of the distance geometry problem in $\mathbb{R}^d$, see Saxe [21].

We formulate the distance geometry problem (1.1) in terms of finding the global minimum of the function

$$f(x) = \sum_{i,j \in \mathcal{S}} w_{i,j} \left(\|x_i - x_j\|^2 - \delta_{i,j}^2 \right)^2, \tag{1.3}$$

where $w_{i,j}$ are positive weights. Clearly, $x \in \mathbb{R}^n$ solves the distance geometry problem if and only if $f(x) = 0$. We could use any global optimization algorithm (see [20], [12], and [5] for global optimization background) in the search for a global minimum of f, but these general algorithms do not take advantage of the structure in the distance geometry problem. Other algorithms used in the solution of distance geometry problems (for example, Hendrickson [10, 11], Havel [9], and Glunt, Hayden, and Raydan [7, 8]) must also rely on general techniques, such as multistarts or simulated annealing, to claim convergence to a global minimizer.

The continuation approach for global optimization hinges on the ability to gradually transform the original function into a smoother function with fewer local minimizers. An optimization algorithm is then applied to the transformed function, tracing their minimizers back to the original function. The idea of transforming a function into a smoother function is appealing; the main approaches include the diffusion equation method of Piela, Kostrowicki, and Scheraga [19], the packet annealing method of Shalloway [24, 23], and the effective energy simulated annealing method of Coleman, Shalloway, and Wu [2, 3]. In the diffusion equation method the transformation can be written (see [13, 14] for details) in the form

$$\frac{1}{(4\pi)^{n/2}\tau^n} \int_{\mathbf{R}^n} f(y) \exp\left(-\frac{\|y - x\|^2}{4\tau} \right) dy, \tag{1.4}$$

where τ is a parameter (time). The smoothing properties of this transformation have been studied by the researchers in Scheraga's group, often in connection with the search for the lowest energy conformation of a molecule (see, for example, [13, 14, 15, 22]). The transformation used in the packet annealing method and in the effective energy simulated

annealing method can be written in the form

$$\frac{1}{\pi^{n/2}|\det\Lambda|^n}\int_{\mathbf{R}^n}\exp\left(-\frac{f(y)}{\kappa_B t}\right)\exp\left(-\|\Lambda^{-1}(y-x)\|^2\right)\,dy, \tag{1.5}$$

where κ_B is the Boltzmann constant, t is a parameter (temperature), and Λ is a nonsingular matrix (the sampling scale). Other transformations used in molecular conformation problems are reviewed by Straub [25]. In this paper we follow the work of Wu [26] by developing the general properties and use of (1.4) in continuation algorithms for the solution of large global optimization problems, since this transformation seems to have the strongest mathematical properties.

We feel that (1.4) is likely to play an important role, not only in the molecular conformation problem, but in the solution of a wide variety of global optimization problems. For this reason Section 2 introduces the term *Gaussian transform* to denote this transformation. We also illustrate the smoothing properties of the general Gaussian transform on a simple two-dimensional problem. This example also provides motivation for the continuation approach.

Section 3 presents some of the more interesting properties of the Gaussian transform. We study, in particular, the computation of the Gaussian transform for the decomposable functions. This is an important class of functions because many of the functions that arise in applications are decomposable. This class of functions was introduced by Wu [26] under the term *generalized multilinear functions*; we are using the term *decomposable* to avoid confusion with the use of multilinear for a function that is linear in each argument.

Our approach for solving the distance geometry problem is outlined in Sections 4 and 5. We compute the Gaussian transform of function (1.3) as a special case of more general results in Section 4, while Section 5 presents the basic ideas behind global continuation algorithms. We concentrate on an approach based on choosing a predetermined sequence of smoothing parameters, since this approach already brings out the power of the continuation algorithm. In future work we plan to address more sophisticated approaches for choosing the smoothing parameters.

In Section 6 we consider a typical distance geometry problem and compare a basic global continuation algorithm with a multistart method for global optimization. We are interested in the solution of problems with a large number of atoms, and thus we performed our numerical testing on the Argonne IBM SP system. This system has 128 nodes, each node an IBM RS/6000-370 with 128 MB of memory. Our main conclusion from the numerical results is that the continuation algorithm finds the solution of the distance geometry problem in all cases but that the multistart method becomes increasingly unreliable and expensive as the number of atoms increases. The reliability of the multistart method drops below 10% for problems with $m \geq 64$ atoms.

2 Continuation for Global Optimization

In the continuation approach for global optimization, the original function is gradually transformed into a smoother function with fewer local minimizers. An optimization algorithm is then applied to the transformed function, tracing the minimizers back to the original function. In this section we define the transformation and provide motivation for the continuation approach.

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

Definition 2.1 *The Gaussian transform* $\langle f \rangle_\lambda$ *of a function* $f : \mathbb{R}^n \mapsto \mathbb{R}$ *is*

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}\lambda^n} \int_{\mathbf{R}^n} f(y) \exp\left(-\frac{\|y-x\|^2}{\lambda^2}\right) dy. \tag{2.1}$$

We are using the term Gaussian transform because we can view $\langle f \rangle_\lambda(x)$ as the expected value of $f(x)$ with respect to the Gaussian density function

$$\rho_\lambda(y) = \frac{1}{\pi^{n/2}\lambda^n} \exp\left(-\frac{\|y\|^2}{\lambda^2}\right). \tag{2.2}$$

The value $\langle f \rangle_\lambda(x)$ of the Gaussian transformation is an average of $f(x)$ in a neighborhood of x, with the relative size of this neighborhood controlled by the parameter λ. The size of the neighborhood decreases as λ decreases so that when $\lambda = 0$, the neighborhood is the center x. We explore additional properties of the Gaussian transformation in the next section.

We illustrate the transformation process with the problem of finding the global maximizer for a function that is a linear combination of four Gaussian functions. The function in the top left corner of Figure 2.1 is of the general form

$$f(x) = \sum_{i=1}^{4} \alpha_i \exp\left(\frac{\|x-x_i\|^2}{\sigma_i^2}\right), \tag{2.3}$$

where $\sigma_i = 0.5$ for $1 \le i \le 4$, $\alpha_1 = 1.5$, and $\alpha_i = 1$ for $i = 2, 3, 4$; the centers x_i are the vertices of the square $[-0.5, 0.5] \times [-0.5, 0.5]$. As can be seen in Figure 2.1, the function has four maximizers in $[-2, 2] \times [-2, 2]$. The Gaussian transforms of (2.3) for three values of λ also appear in Figure 2.1. The top right corner corresponds to $\lambda = 0.2$, and in the bottom row we have $\lambda = 0.3, 0.4$.

Figure 2.1 clearly shows that the original function is gradually transformed into a smoother function with fewer local maximizers, and that the smoothing increases as λ increases. We can view the Gaussian transform of a function as a coarse approximation to the original function, with small and narrow maximizers being removed while the overall

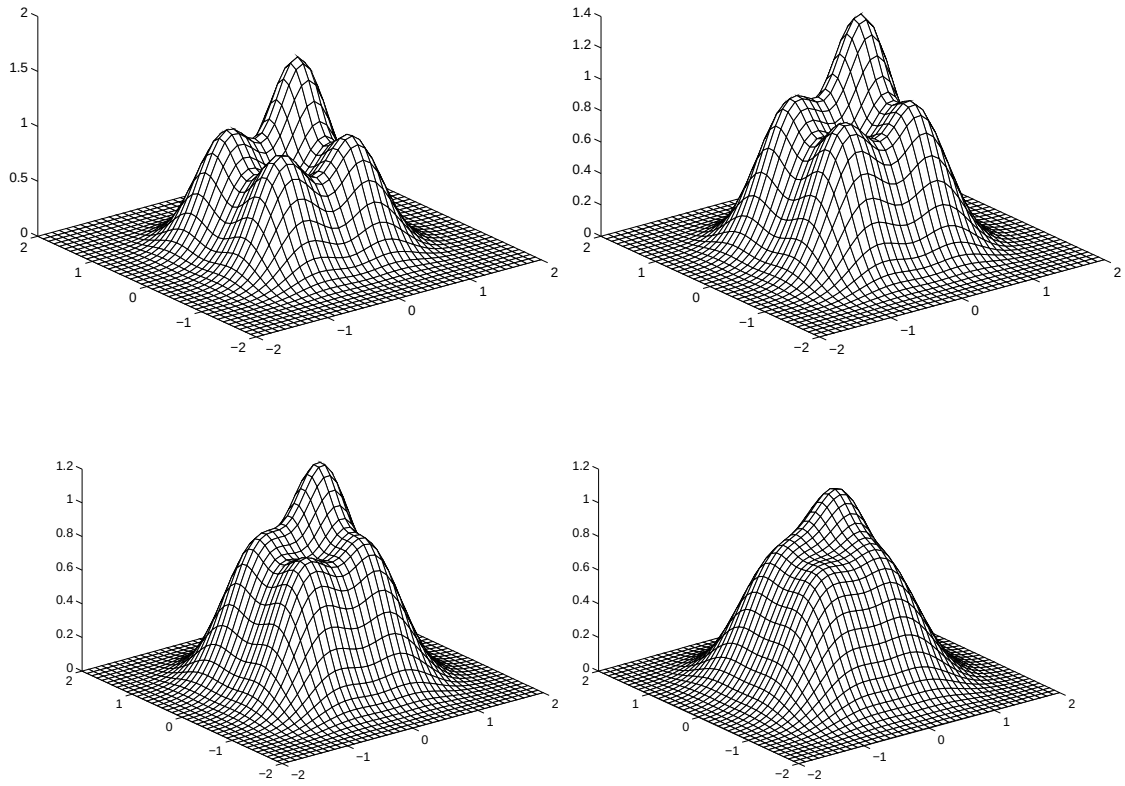

Figure 2.1: The Gaussian transform of a function. The original function ($\lambda = 0$) is in the top left corner, with $\lambda = 0.2$ in the top right corner, $\lambda = 0.3$ in the bottom left corner, and $\lambda = 0.4$ in the bottom right corner.

structure of the function is maintained. This property allows an optimization procedure to skip less interesting local maximizers and to concentrate on regions with average high function values, where a global maximizer is most likely to be located.

Another point that is apparent from Figure 2.1 is that a continuation process based on the Gaussian transform will find the global maximizer. In general, we cannot expect that the continuation process will succeed on an arbitrary function. In particular, the Gaussian transform eliminates tall, narrow hills; hence, if the global maximizer lies in one of these hills, the continuation approach is likely to fail.

3 The Gaussian Transform

We have defined the Gaussian transform for a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ by (2.1). In many cases it is preferable to make the change of variables $y \mapsto x + \lambda u$ in (2.1) to obtain that

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}} \int_{\mathbf{R}^n} f(x + \lambda u) \exp\left(-\|u\|^2\right) \, du. \tag{3.1}$$

In this section we explore some of the properties of this transformation.

The Gaussian transform is defined for a large class of functions. In particular, the transformation is defined if f is continuous almost everywhere and if

$$|f(x)| \leq \beta_1 \exp(\beta_2 \|x\|) \tag{3.2}$$

for positive constants β_1 and β_2. These assumptions guarantee that f is bounded on compact sets, but allow for unbounded f on $\mathbb{R}^n$. In the development that follows, we assume that f satisfies assumptions (3.2).

An important property of this transformation is that $\langle f \rangle_\lambda$ is a linear operator in the sense that

$$\langle \alpha f \rangle_\lambda = \alpha \langle f \rangle_\lambda, \qquad \langle f_1 + f_2 \rangle_\lambda = \langle f_1 \rangle_\lambda + \langle f_2 \rangle_\lambda$$

for any scalar α and functions f_1 and f_2. Also note that the Gaussian transform of the identity function is unity; this result depends on the result

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{\xi^2}{\lambda^2}\right) \, d\xi = \pi^{1/2} \lambda.$$

More generally, if $\mu_1 \leq f(x) \leq \mu_2$ for all $x \in \mathbb{R}^n$, then $\mu_1 \leq \langle f \rangle_\lambda(x) \leq \mu_2$ also holds for all $x \in \mathbb{R}^n$. In particular, this shows that if f is bounded below, then $\langle f \rangle_\lambda$ is also bounded below.

Theorem 3.1 *The Gaussian transform $\langle f \rangle_\lambda$ is a continuous function.*

Proof. The proof is a direct consequence of general results (see, for example, Lang [17, Chapter 13]) on the continuity of functions of the form

$$x \mapsto \int_{\mathbf{R}^n} h(x, y) \, dy,$$

where the mapping h is continuous in x and integrable in y. ∎

Theorem 3.1 helps to support our claim that $\langle f \rangle_\lambda$ is a smoother version of f. Indeed, Theorem 3.1 is a special case of a more general result that establishes $\langle f \rangle_\lambda$ as an infinitely differentiable function. This result can be established by showing that the mapping h defined by

$$h(x, y) = f(y) \rho_\lambda(x - y),$$

where ρ_λ is given by (2.2), is infinitely differentiable with respect to x, and all the derivatives are integrable.

We now show that if f is convex, the Gaussian transform is also a convex function. This property is reassuring because it shows that the transformation does not introduce difficulties if none exist.

Theorem 3.2 *If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, then $\langle f \rangle_\lambda$ is also convex.*

Proof. The result follows from (3.1) because the convexity of f implies that

$$f(\alpha x_1 + (1-\alpha)x_2 + \lambda u) \leq \alpha f(x_1 + \lambda u) + (1-\alpha) f(x_2 + \lambda u), \qquad 0 \leq \alpha \leq 1,$$

for any x_1 and x_2 in $\mathbb{R}^n$. ∎

A serious drawback to the general use of the Gaussian transform for minimization is that computing $\langle f \rangle_\lambda$ for a general function defined on $\mathbb{R}^n$ is not possible because this requires the computation of n-dimensional integrals. However, there is a large class of functions for which the computation of the Gaussian transform is reasonable.

Definition 3.3 *A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is decomposable if f can be written in the form*

$$f(x) = \sum_{k=1}^{m} f_k(x), \qquad f_k(x) = \prod_{j=1}^{n} f_{k,j}(x_j), \tag{3.3}$$

for some set of functions $\{f_{k,j}\}$, where $f_{k,j} : \mathbb{R} \mapsto \mathbb{R}$.

This class of functions was introduced by Wu [26] under the term generalized multilinear functions; we are using decomposable to avoid confusion with the use of multilinear for a function that is linear in each argument.

The decomposable functions are of interest with respect to the Gaussian transform because computing the Gaussian transform of a decomposable function requires the computation of only one-dimensional integrals. Indeed, a computation shows that if f is defined by (3.3), then

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

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

Table 3.1 shows the Gaussian transformation of several elementary functions. We will justify the correctness of the entries later; here we note that the Gaussian transform of any decomposable function with component functions drawn from this table can be calculated explicitly. For example, using these results, we can show that if f is the general quadratic

$$f(x) = \tfrac{1}{2} x^T Q x + c^T x$$

Table 3.1: The Gaussian transformation of elementary functions

$f(x)$	$\langle f\rangle_\lambda(x)$
x	x
x^2	$x^2+\frac{1}{2}\lambda^2$
$\sin(x)$	$\sin(x)\exp(-\frac{1}{4}\lambda^2)$
$\cos(x)$	$\cos(x)\exp(-\frac{1}{4}\lambda^2)$
$\exp(x)$	$\exp(x)\exp(\frac{1}{4}\lambda^2)$

for some $Q\in\mathbb{R}^{n\times n}$ and $c\in\mathbb{R}^n$, then

$$\langle f\rangle_\lambda(x)=\tfrac{1}{2}x^TQx+c^Tx+\tfrac{1}{4}\lambda^2\left(\sum_{i=1}^n q_{i,i}\right). \tag{3.4}$$

In particular, this shows that $\langle f\rangle_\lambda(x)=f(x)$ for linear functions.

Table 3.1 includes only the most commonly occurring functions; there are many other functions with an easily computable Gaussian transform. For example,

$$\langle f\rangle_\lambda(x)=\frac{1}{(\lambda^2+1)^{1/2}}\exp\left(-\frac{x^2}{(\lambda^2+1)}\right)$$

is the Gaussian transform of $f(x)=\exp(-x^2)$.

In addition to quadratic functions, the decomposable functions include the polynomial functions, that is, functions that are linear combinations of terms of the form

$$x_1^{p_1}x_2^{p_2}\cdots x_n^{p_n},$$

for arbitrary integer powers $p_i\geq 0$. The following result is needed to compute $\langle f\rangle_\lambda$ for a polynomial function.

Theorem 3.4 *If $f:\mathbb{R}\mapsto\mathbb{R}$ is the monic polynomial $f(x)=x^k$, then*

$$\langle f\rangle_\lambda(x)=\sum_{l=0}^{\lfloor k/2\rfloor}\left(\frac{k!}{(k-2l)!\,l!}\right)\left(\frac{\lambda}{2}\right)^{2l}x^{k-2l}.$$

Proof. Since f is a polynomial we can expand $f(x+\lambda u)$ in (3.1) and obtain that

$$\langle f\rangle_\lambda(x)=\frac{1}{\pi^{1/2}}\sum_{j=0}^k f^{(j)}(x)\frac{\lambda^j}{j!}\int_{\mathbf{R}}u^j\exp\left(-\|u\|^2\right)\,du,$$

and since the integrals with odd powers vanish by symmetry,

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{1/2}} \sum_{l=0}^{\lfloor k/2 \rfloor} f^{(2l)}(x) \frac{\lambda^{2l}}{(2l)!} \int_{\mathbf{R}} u^{2l} \exp\left(-\|u\|^2\right) \, du.$$

We can complete the proof if we show that

$$\frac{1}{\pi^{1/2}} \int_{\mathbf{R}} u^{2l} \exp\left(-\|u\|^2\right) \, du = \frac{(2l)!}{4^l l!}. \tag{3.5}$$

This identity can be established by defining I_{2l} to be the integral in (3.5) and noting that integration by parts yields

$$I_{2l} = \frac{2l-1}{2} I_{2l-2} = \frac{(2l)(2l-1)}{4l} I_{2l-2}.$$

An induction argument, based on this relationship and using the result $I_0 = 1$, shows that (3.5) holds, and thus completes the proof. ∎

Theorem 3.4 was obtained by Kostrowski and Piela [13], but with a completely different approach. We will elaborate on this point below.

We can extend Theorem 3.4 by noting that if f is analytic, the Taylor series of $f(x + \lambda u)$ as a function of u converges for all λu. Thus we can proceed as in the proof of Theorem 3.4 to obtain

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{1/2}} \sum_{l=0}^{+\infty} f^{(2l)}(x) \frac{\lambda^{2l}}{(2l)!} \int_{\mathbf{R}} u^{2l} \exp\left(-\|u\|^2\right) \, du.$$

Hence, (3.5) shows that

$$\langle f \rangle_\lambda(x) = \sum_{l=0}^{+\infty} \frac{1}{l!} f^{(2l)}(x) \left(\frac{\lambda}{2}\right)^{(2l)}. \tag{3.6}$$

This relationship holds, in particular, for the functions in Table 3.1. A short computation shows that this expression justifies the entries in this table.

Expression (3.6) was used by Piela, Kostrowski, and Scheraga [19] to define the transformation for the diffusion equation method. A disadvantage of this definition is that it requires an analytic f, while (3.1) requires only the integrability of f. On the other hand, as we have noted, this expression is quite useful for determining the Gaussian transform of several important functions. In particular, Kostrowski and Piela [13] obtained Theorem 3.4 with this approach.

The Gaussian transform for functions that are related by a scaling or a translation of the variables can be computed by noting that if

$$f_0(x) = f(\alpha x - x_0)$$

for some scalar α and vector x_0, then

$$\langle f_0 \rangle_\lambda(x) = \langle f \rangle_{\alpha\lambda}(\alpha x - x_0).$$

For example, if $f(x) = \sin(\alpha x)$, then

$$\langle f \rangle_\lambda(x) = \sin(\alpha x) \exp\left(-\tfrac{1}{4}(\alpha\lambda)^2\right).$$

This result suggests that $\langle f \rangle_\lambda$ tends to dampen the high-frequency components in a function, since if α is large, then the exponential term produces a larger damping effect. See Wu [26, Section 4] for a discussion of the effect of the Gaussian transform on the high-frequency components of a general function.

We have defined the Gaussian transform of a real-valued function $f : \mathbb{R}^n \mapsto \mathbb{R}$ by (3.1), but this definition extends immediately to vector-valued functions. This remark is of interest because in addition to transforming the function, we could also transform the gradient and the Hessian of f. We now show that the Gaussian transform of the gradient (Hessian) is the gradient (Hessian) of $\langle f \rangle_\lambda$. This result can be deduced by differentiating under the integral sign in (3.1) to obtain that

$$\nabla \langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}} \int_{\mathbf{R}^n} \nabla f(x + \lambda u) \exp\left(-\|u\|^2\right) du = \langle \nabla f \rangle_\lambda(x), \tag{3.7}$$

which is the desired result for the gradient. If we repeat the process, we obtain that

$$\nabla^2 \langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}} \int_{\mathbf{R}^n} \nabla^2 f(x + \lambda u) \exp\left(-\|u\|^2\right) du = \langle \nabla^2 f \rangle_\lambda(x), \tag{3.8}$$

so that the Gaussian transform of the Hessian matrix is the Hessian of $\langle f \rangle_\lambda$.

We guarantee the validity of differentiating under the integral sign in (3.8) by assuming that $\nabla^2 f$ is continuous almost everywhere and that

$$\|\nabla^2 f(x)\| \leq \gamma_1 \exp(\gamma_2 \|x\|) \tag{3.9}$$

holds for some positive constants γ_1 and γ_2. This result requires a technical lemma.

Lemma 3.5 *If $f : R^n \mapsto \mathbb{R}$ is twice differentiable on $\mathbb{R}^n$ and (3.9) holds for some positive constants γ_1 and γ_2, then*

$$\|\nabla f(x)\| \leq 2\beta_1 \exp(\beta_2 \|x\|), \qquad |f(x)| \leq 3\beta_1 \exp(\beta_2 \|x\|),$$

where $\beta_1 \geq \max\{\gamma_1, \|\nabla f(0)\|, |f(0)|\}$ and $\beta_2 \geq 2 + \gamma_2$.

Proof. The standard estimate

$$\|\nabla f(x) - \nabla f(0)\| \leq \sup_{0 \leq \tau \leq 1} \|\nabla^2 f(\tau x)\| \, \|x\|,$$

together with the estimate $\|x\| \leq \exp(\|x\|)$, implies that

$$\|\nabla f(x)\| \leq \|\nabla f(0)\| + \gamma_1 \exp\left(\gamma_2 \|x\|\right) \|x\| \leq \beta_1 + \beta_1 \exp\left((1 + \gamma_2)\|x\|\right),$$

and thus

$$\|\nabla f(x)\| \leq 2\beta_1 \exp\Big((1+\gamma_2)\|x\|\Big),$$

which is clearly of the desired form. We complete the proof by using this estimate and repeating the above argument, but with ∇f is replaced by f. In this case we obtain

$$|f(x)| \leq |f(0)| + 2\beta_1 \exp\Big((1+\gamma_2)\|x\|\Big)\|x\| \leq \beta_1 + 2\beta_1 \exp\Big((2+\gamma_2)\|x\|\Big),$$

as desired. ∎

We now show that the assumption (3.9) guarantees that (3.7) and (3.8) hold.

Theorem 3.6 *If $f : R^n \mapsto \mathbb{R}$ is twice continuously differentiable almost everywhere on $\mathbb{R}^n$ and (3.9) holds for some positive constants γ_1 and γ_2, then*

$$\nabla\langle f\rangle_\lambda(x) = \langle \nabla f\rangle_\lambda(x), \qquad \nabla^2\langle f\rangle_\lambda(x) = \langle \nabla^2 f\rangle_\lambda(x).$$

Proof. Assumption (3.9) guarantees that the function

$$u \mapsto \nabla^2 f(x+\lambda u)\exp\left(-\|u\|^2\right)$$

is bounded by an integrable function, for any fixed x and λ. The validity of (3.8) now follows from standard results that guarantee differentiation under the integral sign (see, for example, Lang [17, Chapter 13]). Lemma 3.5 shows that the same argument can be used to validate (3.7). ∎

Theorem 3.6 was stated informally by Wu [26]; the above argument supplies the pieces needed to give a formal proof of this result. Theorem 3.6 is of interest from a computational viewpoint because optimization algorithms require the gradient and Hessian of $\langle f\rangle_\lambda$. This result shows that the gradient and Hessian of $\langle f\rangle_\lambda$ are also smooth functions in the sense that they are obtained by transforming the gradient and Hessian, respectively.

In this section we have concentrated on obtaining explicit expressions for the Gaussian transform of various functions. We have also experimented with other approaches. In one of the approaches, the Gaussian transform is approximated by a Gaussian quadrature. This approach hinges on the ability to evaluate Gaussian integrals efficiently with ORTHOPOL (Gautschi [6]). Another approach is based on approximating the function by a decomposable function and using the Gaussian transform of the decomposable function as an approximation to the Gaussian transform of the original function. We plan to pursue these approaches in future work.

4 The Gaussian Transform for the Distance Geometry Problem

Our continuation algorithms for the distance geometry problem are based on the function

$$f(x) = \sum_{i,j\in\mathcal{S}} w_{i,j}\left(\|x_i - x_j\|^2 - \delta_{i,j}^2\right)^2, \tag{4.1}$$

where $w_{i,j}$ are positive weights, and $\delta_{i,j}$ are distances. Computing the Gaussian transform of (4.1) is not difficult because f is decomposable. In fact, f is a polynomial function in the components of x. The development below shows that f has considerable structure and that this structure can be used to simplify the computation for the Gaussian transform.

In the standard formulation of the distance geometry problem, the components $x_i \in \mathbb{R}^3$. We assume that $x_i \in \mathbb{R}^p$ because this assumption does not lead to extra complications. We thus consider the general problem where f is of the form

$$f(x) = \sum_{i,j \in \mathcal{S}} w_{i,j} h_{i,j}(x_i - x_j) \tag{4.2}$$

and $h_{i,j} : \mathbb{R}^p \mapsto \mathbb{R}$ is defined by

$$h_{i,j}(x) = \left(\|x\|^2 - \delta_{i,j}^2 \right)^2. \tag{4.3}$$

The following result shows that computing the Gaussian transform of (4.2) requires only the Gaussian transform on $h_{i,j}$.

Theorem 4.1 *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

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

Proof. Define $Q \in \mathbb{R}^{(n-p) \times n}$ such that

$$R = \frac{1}{\sigma} \begin{pmatrix} P & Q \end{pmatrix}$$

is an orthogonal matrix. By definition,

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}} \int_{\mathbf{R}^n} h(P^T x + \lambda P^T u) \exp\left(-\|u\|^2\right) du,$$

so that if we make the change of variables $u \mapsto Rv$ in (3.1), we obtain

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}} \int_{\mathbf{R}^n} h(P^T x + \lambda P^T R v) \exp\left(-\|v\|^2\right) dv,$$

since R is an orthogonal matrix. Now note that $P^T R = \sigma (I \quad 0)$, and thus the above integral reduces to an integral over $\mathbb{R}^p$, that is,

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{p/2}} \int_{\mathbf{R}^p} h(P^T x + \lambda\sigma v) \exp\left(-\|v\|^2\right) dv = \langle h \rangle_{\sigma\lambda}(P^T x).$$

∎

The application of Theorem 4.1 to the distance geometry problem requires that we specify how the vectors x_i are related to x. Let the i-th component of the vector x_j be the $c(i,j)$ components of x. In other words, $c(i,j)$ specifies how the components of x_j are stored in $x \in \mathbb{R}^n$. Another way of defining $c(i,j)$ is by the relationship

$$[x]_{c(i,j)} = [x_j]_i.$$

With this choice we can set

$$P = \left(e_{c(1,i)} - e_{c(1,j)}, \ldots, e_{c(p,i)} - e_{c(p,j)}\right)$$

and obtain $P^T x = x_i - x_j$. In particular, $P^T P = \sigma^2 I$, where $\sigma^2 = 2$.

As an application of these results, note that Theorem 4.1 implies that

$$\langle f \rangle_\lambda(x) = \langle h \rangle_{\sqrt{2}\lambda}(x_i - x_j).$$

is the Gaussian transform of $f(x) = h(x_i - x_j)$. We can apply this result to the distance geometry problem, where h is given by (4.3), by computing the Gaussian transform of the functions $f_1 : \mathbb{R}^p \mapsto \mathbb{R}$ and $f_2 : \mathbb{R}^p \mapsto \mathbb{R}$ defined by

$$f_1(x) = \|x\|^2, \qquad f_2(x) = \|x\|^4.$$

Since f_1 is a quadratic,

$$\langle f_1 \rangle_\lambda(x) = \|x\|^2 + \tfrac{1}{2} p \lambda^2 \tag{4.4}$$

is just a special case of (3.4). We now claim that Theorem 3.4 shows that

$$\langle f_2 \rangle_\lambda(x) = \|x\|^4 + [3 + (p-1)]\lambda^2 \|x\|^2 + \tfrac{1}{4} p(p+2)\lambda^4. \tag{4.5}$$

We prove (4.5) by noting that

$$\left(\sum_{i=1}^p x_i^2\right)^2 = \sum_{i=1}^p x_i^4 + \sum_{i \neq j}^p \left(x_i^2 x_j^2\right),$$

and thus Theorem 3.4 implies that

$$\langle f_2 \rangle_\lambda(x) = \sum_{i=1}^p \left(x_i^4 + 3\lambda^2 x_i^2 + \tfrac{3}{4}\lambda^4\right) + \sum_{i \neq j}^p \left((x_i^2 + \tfrac{1}{2}\lambda^2)(x_j^2 + \tfrac{1}{2}\lambda^2)\right).$$

Identity (4.5) is now a direct consequence of this expression.

Theorem 4.2 *If $h : \mathbb{R}^p \mapsto \mathbb{R}$ is defined by*

$$h(x) = \left(\|x\|^2 - \delta^2\right)^2,$$

then

$$\langle h \rangle_\lambda(x) = h(x) + [3 + (p-1)]\lambda^2 \|x\|^2 + \tfrac{1}{4} p(p+2)\lambda^4 - p\delta^2\lambda^2.$$

Proof. Since

$$h(x) = f_2(x) - 2\delta^2 f_1(x) + \delta^4,$$

the result follows from (4.4) and (4.5). ∎

The computation of the Gaussian transform for the distance geometry problem now follows from the results that we have obtained.

Theorem 4.3 *If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is defined by (4.2) and (4.3), then*

$$\langle f \rangle_\lambda(x) = f(x) + \sum_{i,j \in \mathcal{S}} \left(2w_{i,j}[3 + (p-1)]\lambda^2 \|x_i - x_j\|^2 \right) + \gamma,$$

where γ is the constant

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

Proof. Recall that we can write f in the form

$$f(x) = \sum_{i,j \in \mathcal{S}} w_{i,j} h_{i,j}(P_{i,j} x),$$

where $P_{i,j}^T P_{i,j} = \sigma^2 I$, with $\sigma^2 = 2$. ∎

Theorem 4.3 shows that the Gaussian transform of the distance geometry function defined by (4.2) and (4.3) can be computed quite easily. Moreover, this result also shows that the gradient and the Hessian matrix of the Gaussian transform are also readily computable at a fractional increase in cost.

We conclude this section by discussing the relationship between Theorem 4.1 and the anisotropic Gaussian transform defined by Wu [26]. Given a nonsingular matrix $\Lambda \in \mathbb{R}^{n \times n}$, the anisotropic Gaussian transform of f is defined by

$$\langle f \rangle_\Lambda(x) = \frac{1}{\pi^{n/2} |\det \Lambda|} \int_{\mathbf{R}^n} f(y) \exp\left(-\|\Lambda^{-1}(y - x)\|^2 \right) \, dy. \tag{4.6}$$

Clearly, this transformation generalizes Definition 2.2, where $\Lambda = \lambda I$.

From a computational viewpoint, the anisotropic transformation is important when Λ is a diagonal matrix, and is closely related to the isotropic transformation when f is a decomposable function. In particular, if f is defined by (3.3) and $\Lambda = \mathrm{diag}(\lambda_j)$, then

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

The following result, a generalization of Theorem 4.1, provides further motivation for the anisotropic transformation.

Theorem 4.4 *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 $P^T P = D^T D$, where D is a diagonal matrix, then

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

Proof. The proof follows that of Theorem 4.1. In this case we define $Q \in \mathbb{R}^{(n-p)\times n}$ such that

$$R = \begin{pmatrix} PD^{-1} & Q \end{pmatrix}$$

is an orthogonal matrix, and obtain that

$$\langle f\rangle_\lambda(x) = \frac{1}{\pi^{p/2}} \int_{\mathbf{R}^p} h(P^T x + \lambda D v) \exp\left(-\|v\|^2\right)\, dv.$$

The result now follows from the definition of the anisotropic transformation because the change of variables $y \mapsto x + Du$ in (4.6) shows that

$$\langle h\rangle_D(x) = \frac{1}{\pi^{n/2}} \int_{\mathbf{R}^n} h(x + Du) \exp\left(-\|u\|^2\right)\, du$$

is the anisotropic transformation of h. ∎

5 Continuation Algorithms

The basic idea behind the continuation approach is to trace a curve $\{x(\lambda) : \lambda \geq 0\}$, where each $x(\lambda)$ is a minimizer of $\langle f\rangle_\lambda$. In the simplest approach we choose a sequence $\{\lambda_k\}$ of smoothing parameters that converges to zero, and compute a minimizer x_k of each $\langle f\rangle_{\lambda_k}$. A more sophisticated approach is to rely on a differential equation to trace the curve. For this approach, we define $h : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$ by

$$h(x, \lambda) = \langle f\rangle_\lambda(x) \tag{5.1}$$

and note that, since $x(\lambda)$ is a stationary point of $\langle f\rangle_\lambda$,

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

We now differentiate with respect to λ to obtain

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

This differential equation, together with an initial value x_0, defines a curve if the coefficient matrix $\partial_{xx} h[x(\lambda), \lambda]$ is nonsingular. In this paper we concentrate on the approach based on

choosing a predetermined sequence of smoothing parameters, since this approach already brings out the power of continuation algorithms.

We wish to analyse the ideal situation where we are able to determine a global minimizer x_k of $\langle f \rangle_{\lambda_k}$ for some sequence $\{\lambda_k\}$ converging to zero. This requires that we show that the function $h : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$ defined by (5.1) is continuous on $\mathbb{R}^n \times \mathbb{R}$. Without loss of generality we show continuity at $(x^*, 0)$. We had previously noted the continuity of h with respect to x and λ; we now establish the joint continuity with respect to (x, λ).

Lemma 5.1 *Assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuous on $\mathbb{R}^n$ and satisfies* (3.2). *If $\{x_k\}$ converges to x^* and $\{\lambda_k\}$ converges to zero, then*

$$\lim_{k \to +\infty} \langle f \rangle_{\lambda_k}(x_k) = f(x^*).$$

Proof. Let B_r be the ball of radius r centered at the origin, and let C_r be the complement of B_r, that is,

$$C_r = \{x \in \mathbb{R}^n : \|x\| > r\}.$$

We first show that for any $\epsilon > 0$ we can choose $r > 0$ and k_0 so that

$$\int_{C_r} |f(x_k + \lambda_k u) - f(x^*)| \exp\left(-\|u\|^2\right) du \leq \epsilon, \qquad k \geq k_0. \tag{5.2}$$

Assumption (3.2) implies that there is a constant $\mu > 0$ such that

$$|f(x_k + \lambda_k u) - f(x^*)| \leq \mu \exp\left(\lambda_k \|u\|\right),$$

and since $\lambda \|u\| \leq \frac{1}{2}\|u\|^2$ for $\lambda \leq \frac{1}{2}$ and $\|u\| \geq 1$,

$$\int_{C_r} |f(x_k + \lambda_k u) - f(x^*)| \exp\left(-\|u\|^2\right) du \leq \mu \int_{C_r} \exp\left(-\tfrac{1}{2}\|u\|^2\right) du$$

if $\lambda_k \leq \frac{1}{2}$ and $r \geq 1$. This estimate proves (5.2) because, if r is sufficiently large, the integral of $\exp\left(-\frac{1}{2}\|u\|^2\right)$ over C_r is arbitrarily small. Now note that the continuity of f at x^* shows that for given r and k_0 we can choose $k_1 \geq k_0$ so that

$$\int_{C_r} |f(x_k + \lambda_k u) - f(x^*)| \exp\left(-\|u\|^2\right) du \leq \epsilon, \qquad k \geq k_1.$$

This estimate and (5.2) imply that

$$|\langle f \rangle_{\lambda_k}(x_k) - f(x^*)| \leq 2\epsilon, \qquad k \geq k_1,$$

which is the desired result. ∎

A variation on Lemma 5.1 would be to show that the gradient and Hessian matrix of h are continuous. The proof of this variation would be entirely similar to that for Lemma 5.1.

Theorem 5.2 *Assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuous on $\mathbb{R}^n$ and satisfies* (3.2). *Let $\{\lambda_k\}$ be any sequence converging to zero. If x_k is a global minimizer of $\langle f \rangle_{\lambda_k}$ and $\{x_k\}$ converges to x^*, then x^* is a global minimizer of f.*

Proof. Since x_k is a global minimizer of $\langle f \rangle_{\lambda_k}$,

$$\langle f \rangle_{\lambda_k}(x_k) \leq \langle f \rangle_{\lambda_k}(x), \qquad x \in \mathbb{R}^n.$$

Lemma 5.1 now implies that $f(x^*) \leq f(x)$ for any $x \in \mathbb{R}^n$. Hence, x^* is a global minimizer of f. ∎

Given λ_k, we need an algorithm to determine a minimizer x_k of $\langle f \rangle_{\lambda_k}$. A trust region version of Newton's method based on the work of Moré and Sorensen [18] is an attractive choice because it has strong global and local convergence properties.

At each iteration of a trust region Newton method for the minimization of $f : \mathbb{R}^n \mapsto \mathbb{R}$, we have an iterate x_k, a bound Δ_k, a scaling matrix D_k, and a quadratic model $q_k : \mathbb{R}^n \mapsto \mathbb{R}$ of the possible reduction $f(x_k + w) - f(x_k)$ for $\|D_k w\| \leq \Delta_k$. The developments in Section 4 show that the gradient and Hessian matrix can be easily obtained for the distance geometry problem. Thus

$$q_k(w) = \nabla f(x_k)^T w + \tfrac{1}{2} w^T \nabla^2 f(x_k) w$$

is our choice for the quadratic model.

An important ingredient in a trust region method is the choice of step s_k. In general s_k is an approximate solution to the trust region subproblem

$$\min \{q_k(w) : \|D_k w\| \leq \Delta_k\}$$

with $q_k(s_k) < 0$. We use the algorithm described by Moré and Sorensen [18] because it provides an approximate global solution to the subproblem. In particular, if x_k is a saddle point so that $\nabla f(x_k) = 0$ and $\nabla^2 f(x_k)$ is indefinite, we still have $q_k(s_k) < 0$.

Given the step s_k, the test for acceptance of the trial point $x_k + s_k$ depends on a parameter $\eta_0 > 0$. The following algorithm summarizes the main computational steps:

For $k = 0, 1, \ldots,$ maxiter
- Compute the quadratic model q_k.
- Compute a scaling matrix D_k.
- Compute an approximate solution s_k to the trust region subproblem.
- Compute the ratio ρ_k of actual to predicted reduction.
- Set $x_{k+1} = x_k + s_k$ if $\rho_k \geq \eta_0$; otherwise set $x_{k+1} = x_k$. Update Δ_k.

Given a step s_k such that $\|D_k s_k\| \leq \Delta_k$ and $q_k(s_k) < 0$, the rules for updating the iterate x_k and the bound Δ_k depend on the ratio

$$\rho_k = \frac{f(x_k + s_k) - f(x_k)}{q_k(s_k)}$$

of the actual reduction in the function to the predicted reduction in the model. See, for example, Moré and Sorensen [18] for details on these rules.

The trust region method outlined above is attractive for the distance geometry problem provided the number of molecules m is moderate, say $m \leq 50$. For larger problems we can still use the trust region method provided the set $\mathcal{S}$ in (4.1) is sparse and the computation of the step s_k makes use of sparsity. We plan to address this case in future work.

6 Numerical Results

Consider a molecule with $m = s^3$ atoms located in the three-dimensional lattice

$$\{(i_1, i_2, i_3) : 0 \leq i_1 < s,\ 0 \leq i_2 < s,\ 0 \leq i_3 < s\}$$

for some integer $s \geq 1$. Figure 6.1 shows a molecule with 64 atoms ($s = 4$). We specify an ordering for the atoms in this molecule by letting atom i be the atom at position (i_1, i_2, i_3), where

$$i = 1 + i_1 + si_2 + s^2 i_3.$$

Given a subset $\mathcal{S}$ of the pairwise distances $\delta_{i,j}$ between atoms i and j, we consider the distance geometry problem

$$\|x_i - x_j\| = \delta_{i,j}, \qquad (i,j) \in \mathcal{S}, \tag{6.1}$$

where the set $\mathcal{S}$ is defined in terms of an integer r by

$$\mathcal{S} = \{(i,j) : |i - j| \leq r\}. \tag{6.2}$$

With this definition the set $\mathcal{S}$ is sparse in the sense that it contains only rm pairs, out of a possible m^2 pairs. Figure 6.2 shows an 8-atom problem defined by a sparse $\mathcal{S}$ with $r = 3$.

The construction of our model problem is realistic in the sense that distance constraints are imposed only on nearby atoms. A computation shows that for the model problem

$$\delta_{i,j} \leq \left(2(s-1)^2 + 1\right)^{\frac{1}{2}} \leq \sqrt{2}\, s.$$

Our construction shows that the distance geometry problem defined by (6.1) and (6.2) always has at least one solution.

We attack the distance geometry problem by using the global continuation approach to obtain a global minimum of the function

$$f(x) = \sum_{(i,j)\in\mathcal{S}} (\|x_i - x_j\|^2 - \delta_{i,j}^2)^2, \tag{6.3}$$

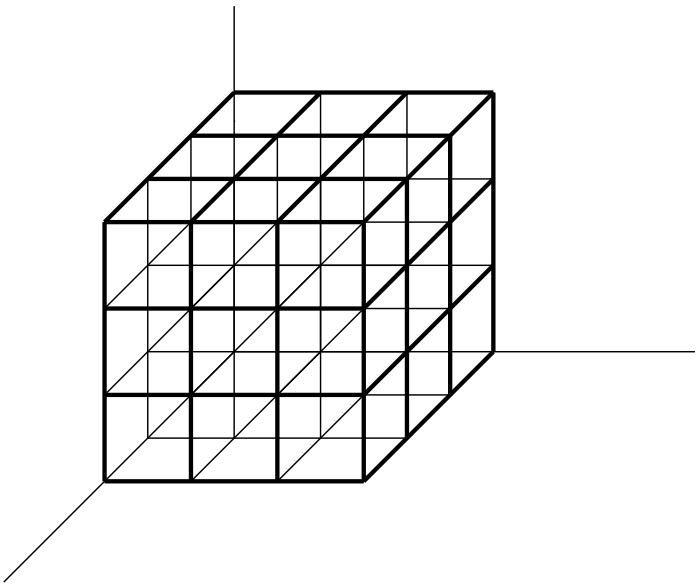

Figure 6.1: An example lattice structure of 64 atoms

where $\delta_{i,j}$ is the distance between atoms i and j in the lattice. We need the Gaussian transform of f and, for the trust region Newton method, the gradient and Hessian matrix of the transform. Theorem 4.3 shows that

$$\langle f \rangle_\lambda(x) = \sum_{(i,j)\in\mathcal{S}} \left((\|x_i - x_j\|^2 - \delta_{i,j}^2)^2 + 10\lambda^2 \|x_i - x_j\|^2 \right) + \gamma \tag{6.4}$$

is the Gaussian transform of f, where γ is a constant. The gradient and Hessian matrix can be obtained from this expression.

For this problem, to determine λ so that $\langle f \rangle_\lambda$ is convex, we study the dependence of $\langle f \rangle_\lambda$ on λ in terms of the function

$$h(r) = \left(r^2 - \delta^2 \right)^2 + 10\lambda^2 r^2.$$

Note that if h is convex, then $x \mapsto h(\|x_i - x_j\|)$ is convex, and thus $\langle f \rangle_\lambda$ is also convex. If we choose λ so that h is convex, then

$$h''(r) = 12r^2 - 8\delta^2 + 20\lambda^2$$

shows that we must have $\lambda \geq (\frac{2}{5})^{1/2}\delta \approx 0.63\delta$. In particular, for a fixed value of λ, terms with a smaller δ have smaller regions of nonconvexity than those with larger values of δ. Also note that if

$$\lambda \geq \left(\frac{2}{5} \right)^{1/2} \max \left\{ \delta_{i,j} : (i,j) \in \mathcal{S} \right\},$$

then $\langle f \rangle_\lambda$ is convex.

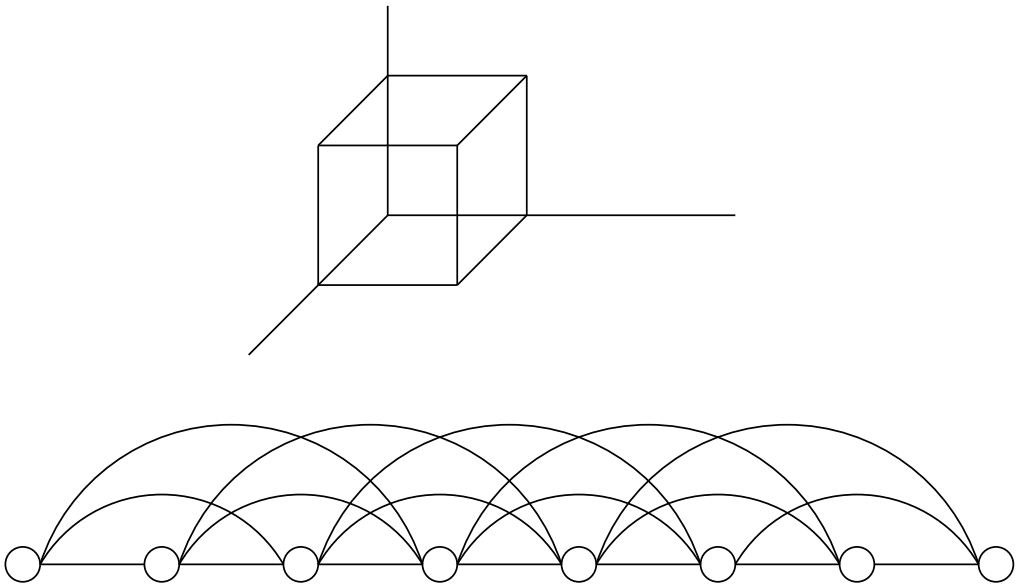

Figure 6.2: An example lattice structure with sparse distance constraints

In this paper we have shown that the continuation method has strong theoretical properties. We now use numerical results to show that the continuation method is superior to the multistart approach, a standard procedure for finding the global minimizer of f.

We are interested in the solution of problems with a large number of atoms, and thus we performed our numerical testing on the Argonne IBM SP system. This system has 128 nodes, each node an IBM RS/6000-370 with 128 MB of memory.

In the multistart method we choose a random starting point x_s and use the trust region method from this starting point to determine a local minimizer x_s^*. If x_s^* satisfies

$$\left| \|x_i - x_j\| - \delta_{i,j} \right| \leq \epsilon, \qquad (i,j) \in \mathcal{S}, \tag{6.5}$$

for some tolerance ϵ, then x_s^* is declared to be a solution to the distance geometry problem (6.1), and we terminate the multistart method. If x_s^* does not satisfy (6.5), we repeat the procedure with another starting point. The multistart method fails if (6.5) is not satisfied after trying ten starting points.

The global continuation method that we use is similar to the multistart method, except that the continuation algorithm of Section 5 is used to determine a local minimizer x_s^* of f. We start the continuation algorithm with the random starting point x_s and $\lambda_0 > 0$. We compute p major iterations, where p is the number of continuation steps. The k-th major iterate x_k is computed by applying a trust region algorithm, with x_{k-1} as a starting point, to the transformed function $\langle f \rangle_{\lambda_k}$, where

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

Table 6.1: Performance of the multistart and continuation methods ($x_s \in \mathbf{rand}(B)$)

		Multistart		Continuation	
m	r	$nfev$	$ngev$	$nfev$	$ngev$
27	9	573	472	255	216
64	16	F1211	1009	886	710
125	25	1810	1461	390	304
216	36	F3397	2782	550	421

Table 6.2: Performance of the multistart and continuation methods ($x_s \in 2\,\mathbf{rand}(B)$)

		Multistart		Continuation	
m	r	$nfev$	$ngev$	$nfev$	$ngev$
27	9	273	229	221	188
64	16	1102	917	863	698
125	25	1600	1324	410	322
216	36	F3416	2802	446	337

Since $\lambda_p = 0$, the final major iterate x_p is a local minimizer of f, so we set $x_s^* = x_p$.

In Tables 6.1 and 6.2 we present the results obtained by the global continuation method and the multistart method on two sets of starting points. The number of molecules in these tables are of the form $m = s^3$ for $3 \leq s \leq 6$. The parameter r in (6.2) was set to $r = s^2$.

Since the solution of the distance geometry problems defined by (6.1) and (6.2) lie in

$$B = \{x \in \mathbb{R}^n : 0 \leq x_i \leq s - 1\},$$

it is reasonable to choose the starting points randomly in B by setting each component of the starting point to a random number in $(0, s-1)$. These results appear in Table 6.1. Similarly, for the results shown in Table 6.2 we choose the starting point randomly in $2B$.

For these results we used $\lambda_0 = 0.5$ and $p = 10$ continuation steps. (Later we consider how the performance of the continuation method depends on λ_0 and p.)

Performance is measured in terms of the number of function and gradient evaluations, `nfev` and `ngev`, used to find a global minimizer. The results marked by F are the cases where no global minimizer was found after trying 10 starting points.

We have not included execution times in Tables 6.1 and 6.2 because the distance geometry problems under consideration give rise to sparse minimization problems, but the

Table 6.3: Probability of success of the multistart and continuation methods

m	r	Multistart	Continuation	Multistart	Continuation
27	9	10%	100%	60%	100%
64	16	0%	70%	10%	50%
125	25	10%	100%	10%	100%
216	36	0%	100%	0%	100%
		$x_s \in \mathbf{rand}(B)$		$x_s \in 2\,\mathbf{rand}(B)$	

algorithm that we have used does not take advantage of sparsity. Our concern in this paper is mainly with the ability of the continuation method to solve these problems with a reasonable number of function and gradient evaluations. In future work we will consider problems with more atoms and the use of algorithms that take advantage of sparsity.

These results show that the continuation method finds a global minimizer in all cases, and with fewer function and gradient evaluations than the multistart method. Moreover, the performance of the continuation method seems to be relatively insensitive to the choice of starting point. The multistart method, on the other hand, requires a large number of function and gradient evaluations to determine a global minimizer, and is unable to find a global minimizer for problems with $m = 216$ atoms. Also note that the performance of the multistart method seems to be sensitive to the choice of starting point.

The reliability of the continuation and multistart methods can be measured by the probability of success of these methods, that is, the percentage of successful runs (the global minimizer is found) in all ten starting points. The results in Table 6.3 clearly show that the multistart method had little success in finding a global minimizer, especially for problems with $m \geq 64$ atoms. However, the continuation method succeeded 100% in most of the cases. Even for $m = 64$, the probability of success is much higher for the continuation method.

One might wonder why the continuation method was not able to find the global minimizer for $m = 64$ in all ten runs. A simple answer to this question is that the initial $\lambda_0 = 0.5$ value was too small for smoothing the function in this problem. Therefore, we repeated the runs for the problem with $m = 64$ atoms, but with $\lambda_0 = 1$ and $p = 20$. The continuation method then found the global minimizer for all ten starting points.

The results in Table 6.4 compare the average performance of the multistart and the continuation method when $\lambda_0 = 1$ and $p = 20$. When $m = 64$ and $x_s \in \mathbf{rand}(B)$, the multistart method fails in all cases, so the results in Table 6.4 measure the effort required to find a local minimizer. In contrast, the continuation method succeeds in all cases, so the

Table 6.4: Average performance for the multistart and continuation method ($\lambda_0 = 1, p = 20$)

		Multistart		Continuation		Multistart		Continuation	
m	r	$nfev$	$ngev$	$nfev$	$ngev$	$nfev$	$ngev$	$nfev$	$ngev$
27	9	61.2	50.5	251.1	211.1	57.4	98.7	240.9	200.9
64	16	121.1	100.9	267.9	212.4	118.3	98.7	272.2	217.0
125	25	241.2	197.3	328.4	249.2	212.1	176.5	344.9	265.5
216	36	339.7	278.2	446.9	340.8	341.6	280.2	472.6	361.7
		$x_s \in \mathbf{rand}(B)$				$x_s \in 2\,\mathbf{rand}(B)$			

results measure the effort required to find a global minimizer. This is interesting because, in general, we expect the effort required to find a global minimizer to be much larger than the effort needed to find a local minimizer. A similar conclusion is reached when $m = 64$ and $x_s \in 2\,\mathbf{rand}(B)$, since in this case the multistart method only succeeds in one case. When $m = 216$, the effort (measured by the number of function and gradient evaluations) required to find a global minimizer is less than 30% more than the effort required to find a local minimizer.

Acknowledgment. Special thanks go to Paul Bash, Danny Ripoll, and Tamar Schlick for the time that they spent talking to us about problems in biology and the relevant literature. Thanks also go to Tom Coleman and Bruce Hendrickson for sharing their experiences with the distance geometry problem.

References

[1] 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.

[2] 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.

[3] ——, *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.

[4] G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, 1988.

[5] C. Floudas and P. Pardalos, eds., *Recent Advances in Global Optimization*, Princeton University Press, 1992.

[6] 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.

[7] W. Glunt, T. L. Hayden, and M. Raydan, *Molecular conformation from distance matrices*, J. Comp. Chem., 14 (1993), pp. 114–120.

[8] ——, *Preconditioners for distance matrix algorithms*, J. Comp. Chem., 15 (1994), pp. 227–232.

[9] 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.

[10] B. A. Hendrickson, *The molecule problem: Determining conformation from pairwise distances*, PhD thesis, Cornell University, Ithaca, New York, 1991.

[11] ——, *The molecule problem: Exploiting structure in global optimization*, SIAM J. Optimization, (1995). To Appear.

[12] R. Horst and H. Tuy, *Global Optimization*, Springer-Verlag, 1990.

[13] J. Kostrowicki and L. Piela, *Diffusion equation method of global minimization: Performance for standard functions*, J. Optim. Theory Appl., 69 (1991), pp. 269–284.

[14] 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.

[15] 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.

[16] 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.

[17] S. Lang, *Real Analysis*, Addison-Wesley, second ed., 1983.

[18] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.

[19] 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.

[20] A. H. G. Rinnooy Kan and G. T. Timmer, *Global optimization*, in Optimization, G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, eds., North-Holland, 1989, pp. 631–662.

[21] 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.

[22] 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.

[23] D. Shalloway, *Application of the renormalization group to deterministic global minimization of molecular conformation energy functions*, J. Global Optim., 2 (1992), pp. 281–311.

[24] D. Shalloway, *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.

[25] J. E. Straub, *Optimization techniques with applications to proteins*, preprint, Boston University, Department of Chemistry, Boston, Massachusetts, 1994.

[26] 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.