

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

**SMOOTHING TECHNIQUES FOR MACROMOLECULAR GLOBAL
OPTIMIZATION**

Jorge J. Moré and Zhijun Wu

Mathematics and Computer Science Division

Preprint MCS-P542-0995

September 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.

ABSTRACT

We study global optimization problems that arise in macromolecular modeling, and the solution of these problems via continuation and smoothing. Our results unify and extend the theory associated with the use of the Gaussian transform for smoothing. We show that the Gaussian transform can be viewed as a special case of a generalized transform and that these generalized transforms share many of the properties of the Gaussian transform. We also show that the smoothing behavior of the generalized transform can be studied in terms of the Fourier transform and that these results indicate that the Gaussian transform has superior smoothing properties.

SMOOTHING TECHNIQUES FOR MACROMOLECULAR GLOBAL OPTIMIZATION

Jorge J. Moré and Zhijun Wu

1 Introduction to Macromolecular Global Optimization

Macromolecular modeling gives rise to a wide variety of global optimization problems with interesting features. These problems usually require the determination of the global minimum of a function with $n = 3m$ variables, where m is the number of atoms in the molecule. A typical problem requires the determination of a molecular structure such that the positions of the atoms x_1, \dots, x_m in \mathbb{R}^3 satisfy certain constraints on the structure or such that a given potential energy function is minimized. There is also interest in determining structures that nearly satisfy these constraints or that have nearly minimal potential value.

The problem of determining a structure that satisfies geometric constraints on the structure falls under the general area of distance geometry. Distance data between atoms and other geometric constraints (for example, angle constraints) can be obtained from nuclear magnetic resonance (NMR) data, but in most cases only partial information is available. In particular, for distance data only a fraction of the distances $\delta_{i,j}$ between the (i, j) pair of atoms is available for a subset \mathcal{S} of the atom pairs. Given this information, we seek positions x_1, \dots, x_m in \mathbb{R}^3 of the atoms in the molecule such that

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

Since the data obtained from NMR is usually inaccurate, distance geometry problems that arise in the determination of protein structure are usually associated with the more general problem of finding positions x_1, \dots, x_m in \mathbb{R}^3 such that

$$l_{i,j} \leq \|x_i - x_j\| \leq u_{i,j}, \quad (i, j) \in \mathcal{S}, \quad (1.2)$$

where $l_{i,j}$ and $u_{i,j}$ are lower and upper bounds on the distance constraints, respectively.

Problems (1.1) and (1.2) can be formulated in terms of finding the global minimum of a function that measures the deviation of the structure from the constraints. We formulate problem (1.1) in terms of finding the global minimum of the function

$$f(x) = \sum_{i,j \in \mathcal{S}} f_{i,j}(x_i - x_j), \quad (1.3)$$

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.

where

$$f_{i,j}(x) = \left(\|x\|^2 - \delta_{i,j}^2 \right)^2. \quad (1.4)$$

Similarly, we formulate problem (1.2) in terms of finding the global minimum of (1.3) where

$$f_{i,j}(x) = \min^2 \left\{ \frac{\|x\|^2 - l_{i,j}^2}{l_{i,j}^2}, 0 \right\} + \max^2 \left\{ \frac{\|x\|^2 - u_{i,j}^2}{u_{i,j}^2}, 0 \right\}. \quad (1.5)$$

For both of these problems, $x = \{x_1, \dots, x_m\}$ solves problem (1.2) if and only if x is a global minimizer of f and $f(x) = 0$ (see Crippen and Havel [4]).

In general, distance geometry calculations impose additional constraints on the structure. For additional information on distance geometry, see Crippen and Havel [4], Havel [11], Kuntz, Thomason, and Oshiro [16], and Brünger and Nilges [1].

In macromolecular modeling we are also interested in determining structures with minimal or nearly minimal potential energy. Various potential functions arise in applications. We are interested, in particular, in the determination of positions x_1, \dots, x_m in \mathbb{R}^3 of the atoms so that the potential energy

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

is minimized. For information and references to this problem, see Hasse and Schiffer [10], Rafac, Schiffer, Hangst, Dubin, and Wales [21], and Schiffer [25].

The problems that we have introduced can be described in terms of finding the global minimum of a function of the form

$$f(x) = \sum_{i,j \in \mathcal{S}} f_{i,j}(x_i - x_j) + \sum_{i=1}^m f_i(x_i), \quad (1.6)$$

where

$$f_{i,j}(x) = h_{i,j}(\|x\|), \quad f_i(x) = h_i(\|x\|) \quad (1.7)$$

for some functions $h_{i,j} : \mathbb{R} \mapsto \mathbb{R}$ and $h_i : \mathbb{R} \mapsto \mathbb{R}$. Our approach for determining the global minimum of functions of the general form (1.6,1.7) is based on smoothing and continuation. The smoothing process transforms a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ into a smoother function $\langle f \rangle_\lambda$ with fewer local minimizers. The parameter λ controls the degree of smoothing; the original function is obtained if $\lambda = 0$, while smoother functions are obtained as λ increases. Given the smooth function $\langle f \rangle_\lambda$, we use continuation on λ to trace the minimizers of $\langle f \rangle_\lambda$ back to the original function.

In this paper we unify and extend the work of Wu [31] and Moré and Wu [18, 19] on using the Gaussian transform $\langle f \rangle_\lambda$ for smoothing. Our results in Sections 2 and 3 show that the Gaussian transform can be viewed as a special case of a generalized transform and that these generalized transforms share many of the properties of the Gaussian transform. Of

interest are the results in Section 3 on the smoothing behavior of the generalized transform in terms of the Fourier transform, since these results indicate that the Gaussian transform has superior smoothing properties.

Sections 5 and 6 complete this paper with a discussion of techniques for computing the generalized transform. Section 5 introduces the decomposable functions and shows how the generalized transform of a decomposable function can be expressed in terms of one-dimensional integrals, which can be either evaluated analytically or, at worst, approximated by Gaussian quadratures. Section 6 outlines several techniques that can be used to compute the Gaussian transform of any function defined by (1.6,1.7). We show that the Gaussian transform can be expressed in terms of special functions or one-dimensional integrals. These techniques are only applicable to the Gaussian transform, and thus show that the Gaussian transform plays a special role in problems that arise in macromolecular modelling.

2 Smoothing

The value $\langle f \rangle_\lambda$ of the transformed function at a point $x \in \mathbb{R}^n$ is generally obtained by computing weighted averages of f in a neighborhood centered at x . The weights are determined by a probability density function, that is, a nonnegative function $\rho : \mathbb{R}^n \mapsto \mathbb{R}$ such that

$$\int_{\mathbb{R}^n} \rho(x) dx = 1.$$

We could use any density function, but the use of the Gaussian density function

$$\rho(y) = \frac{1}{\pi^{n/2}} \exp(-\|y\|^2), \quad (2.1)$$

is natural. We also want the weight of the density function concentrated on the origin, so we use the Gaussian density function

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

with mean zero and variance $\frac{1}{2}\lambda^2$.

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

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2} \lambda^n} \int_{\mathbb{R}^n} f(y) \exp\left(-\frac{\|y-x\|^2}{\lambda^2}\right) dy. \quad (2.3)$$

This definition shows that the value of the transformed function $\langle f \rangle_\lambda$ at a point $x \in \mathbb{R}^n$ is a weighted average of the values of f in a neighborhood of x . This observation can also be made by noting that the change of variables $y = x + \lambda u$ in (2.3) shows that the Gaussian transform can be written in the form

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} f(x + \lambda u) \exp(-\|u\|^2) du. \quad (2.4)$$

Finally, one can also think of the Gaussian transform as the convolution of f with the Gaussian density function ρ_λ . This last observation comes into play when we analyze the smoothing properties of the Gaussian transform.

We have defined the Gaussian transform for a vector-valued mapping $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ because we are also interested in the more general case. In particular, this extended definition will come up when discussing the smoothing of gradients and Hessian matrices.

The idea of transforming a function into a smoother function has been used extensively in macromolecular modeling. Straub [29] reviews work in this area. The diffusion equation method of Piela, Kostrowicki, and Scheraga [20] is central to our work since the idea of the Gaussian transform underlies the diffusion equation method. Other approaches include the packet annealing method of Shalloway [27, 26] and the effective energy simulated annealing method of Coleman, Shalloway, and Wu [2, 3].

We generalize the definition of the Gaussian transform by replacing the Gaussian density function with another density function $\rho : \mathbb{R}^n \mapsto \mathbb{R}$.

Definition 2.2 *Given a density function $\rho : \mathbb{R}^n \mapsto \mathbb{R}$, the generalized transform $\langle\langle f \rangle\rangle_\lambda$ of a function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is*

$$\langle\langle f \rangle\rangle_\lambda(x) = \frac{1}{\lambda^n} \int_{\mathbb{R}^n} f(y) \rho\left(\frac{x-y}{\lambda}\right) dy. \quad (2.5)$$

Transformations of the general form (2.5) have been used in stochastic optimization, often in the equivalent formulation

$$\langle\langle f \rangle\rangle_\lambda(x) = \int_{\mathbb{R}^n} f(x - \lambda v) \rho(v) dv. \quad (2.6)$$

Any density function can be used in (2.5) but from a computational viewpoint, the Gaussian density function (2.1) and the uniform density function

$$\rho(x) = \frac{1}{2}, \quad \|x\|_\infty \leq 1, \quad (2.7)$$

are of special interest.

The earliest reference to transformations of the form (2.5) in stochastic optimization seems to be due to Katkovnik (see, for example, Katkovnik and Kulchitskii [12]). References and generalizations of this work can be found in the work of Rubinstein [22, 23, 24], where the generalized transformation (2.5) is called a *smoothed functional*. Related work and other references can be found in Kreimer and Rubinstein [15] and Ermoliev, Norkin, and Wets [6]. Transformations of the form (2.5) also arise in the theory of distributions, but in this work the density ρ is usually infinitely differentiable and of compact support.

The emphasis of the work on stochastic optimization has been on the approximation of non-smooth functions. There has been little emphasis on computational issues. In particular, the transformation (2.5) is never computed explicitly; instead, a Monte Carlo method is used to approximate the transformation. As we shall see, the emphasis of the work on global optimization is quite different.

3 Properties of the Generalized Transform

The Gaussian transform has many interesting properties, with most of these properties being shared by the generalized transform $\langle\langle f \rangle\rangle_\lambda$. We first consider the questions of existence of the transformations.

The Gaussian transform is defined if f is continuous almost everywhere, and if

$$|f(x)| \leq \beta_1 \exp(\beta_2 \|x\|) \quad (3.1)$$

for positive constants β_1 and β_2 . Existence follows from (2.4), since (3.1) implies that

$$\left| f(x + \lambda v) \exp(-\|v\|^2) \right| \leq \beta_1 \exp(\beta_2 \|x\|) \exp(\beta_2 \lambda \|v\| - \|v\|^2).$$

The situation for the generalized transform is more delicate. The generalized transform $\langle\langle f \rangle\rangle_\lambda$ is defined if we assume that the density function ρ in (2.5) is continuous almost everywhere and has compact support. We can also admit density functions with infinite support if we restrict the class of functions f and density functions ρ . For example, if we assume that

$$|f(x)| \leq \beta_1 + \beta_2 \|x\|^p$$

holds for positive constants β_1 and β_2 and exponent p , then the generalized transform $\langle\langle f \rangle\rangle_\lambda$ is defined if

$$\int_{\mathbb{R}^n} \|v\|^k \rho(v) dv, \quad k = 0, 1, \dots, p$$

exists. In this section we assume that f satisfies assumption (3.1) whenever we are discussing the Gaussian transform $\langle f \rangle_\lambda$ or the generalized transform $\langle\langle f \rangle\rangle_\lambda$ for a density function with compact support. The modifications needed to handle arbitrary density functions should be clear from this discussion.

The generalized transform is a linear operator in the vector space of functions that are continuous almost everywhere and satisfy (3.1), since

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

for any scalar α and functions f_1 and f_2 . The generalized transform $\langle\langle f \rangle\rangle_\lambda$ is also an isotone operator, since it preserves the standard order relation between functions, that is,

$$f_1 \leq f_2 \quad \text{implies that} \quad \langle\langle f_1 \rangle\rangle_\lambda \leq \langle\langle f_2 \rangle\rangle_\lambda.$$

In particular,

$$\min \{f(y) : \mathbb{R}^n\} \leq \langle\langle f \rangle\rangle_\lambda(x) \leq \max \{f(y) : \mathbb{R}^n\}.$$

These properties are direct consequences of the definition of the generalized transform $\langle\langle f \rangle\rangle_\lambda$.

The differentiability properties of the generalized transform $\langle\langle f \rangle\rangle_\lambda$ follow from general results (see, for example, Lang [17, Chapter 13]) on the differentiability of functions of the form

$$g(x) = \int_{\mathbb{R}^n} h(x, y) dy,$$

where the mapping $h : \mathbb{R}^n \mapsto \mathbb{R}$ is integrable in y . If $\partial_x h$ is continuous almost everywhere in an open set of the form $B \times \mathbb{R}^n$, and

$$|\partial_x h(x, y)| \leq h_B(y), \quad (x, y) \in B \times \mathbb{R}^n, \quad (3.2)$$

for some integrable function $h_B : \mathbb{R}^n \mapsto \mathbb{R}$, then g is differentiable and

$$g'(x) = \int_{\mathbb{R}^n} \partial_x h(x, y) dy.$$

This result can be applied, in particular, to the Gaussian transform.

Theorem 3.1 *The Gaussian transform $\langle f \rangle_\lambda$ is infinitely differentiable.*

Proof. Define

$$h(x, y) = f(y) \rho\left(\frac{x - y}{\lambda}\right),$$

and note that

$$\partial_x h(x, y) = \frac{1}{\lambda} f(y) \rho'\left(\frac{x - y}{\lambda}\right).$$

A calculation based on (3.1) shows that (3.2) holds, and thus $\langle f \rangle_\lambda$ is differentiable. The argument can be repeated to show that $\langle f \rangle_\lambda$ is infinitely differentiable. ■

Theorem 3.1 does not hold for the generalized transform $\langle\langle f \rangle\rangle_\lambda$ unless we make additional assumptions on f or on the density function ρ . Consider, for example, the uniform density function

$$\rho(s) = \frac{1}{2}, \quad |s| \leq 1.$$

In this case the generalized density function reduces to

$$\langle\langle f \rangle\rangle_\lambda(x) = \frac{1}{2} \int_{-1}^1 f(x - \lambda s) ds,$$

and thus

$$\langle\langle f \rangle\rangle'_\lambda(x) = \frac{1}{2\lambda} (f(x + \lambda) - f(x - \lambda)).$$

This calculation shows that $\langle\langle f \rangle\rangle_\lambda$ is differentiable but that if f is not continuous, then $\langle\langle f \rangle\rangle_\lambda$ is not continuously differentiable.

The functions that arise in our work on macromolecular modeling are usually twice differentiable, and in this case we can show that the generalized transform $\langle\langle f \rangle\rangle_\lambda$ is also twice differentiable.

Theorem 3.2 *If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is twice continuously differentiable almost everywhere on \mathbb{R}^n and*

$$\|\nabla^2 f(x)\| \leq \gamma_1 \exp(\gamma_2 \|x\|) \quad (3.3)$$

for some positive constants γ_1 and γ_2 , then

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

An informal proof of this result can be obtained by differentiating under the integral sign in (2.4) to obtain that

$$\nabla \langle\langle f \rangle\rangle_\lambda(x) = \int_{\mathbb{R}^n} \nabla f(x - \lambda u) \rho(u) du = \langle\langle \nabla f \rangle\rangle_\lambda(x),$$

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

$$\nabla^2 \langle\langle f \rangle\rangle_\lambda(x) = \int_{\mathbb{R}^n} \nabla^2 f(x - \lambda u) \rho(u) du = \langle\langle \nabla^2 f \rangle\rangle_\lambda(x),$$

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

Theorem 3.2 was stated informally by Wu [31] for the Gaussian density function; a formal proof under assumption (3.3) appears in Moré and Wu [18]. The proof for the generalized transform follows the arguments used by Moré and Wu [18]; the only tricky part in the proof is to show that assumption (3.3) guarantees that we can differentiate under the integral sign.

Theorem 3.2 is of interest from a computational viewpoint because optimization algorithms require the gradient and Hessian of $\langle\langle f \rangle\rangle_\lambda$. This result shows that the gradient and Hessian of $\langle\langle f \rangle\rangle_\lambda$ are also smooth functions in the sense that they are obtained by transforming the gradient and Hessian matrix, respectively.

4 Smoothing Properties of the Generalized Transform

One of the main attractions of the generalized transform is the ability to reduce the high-frequency components of the function. We quantify this statement in terms of the Fourier transform

$$\hat{f}(w) = \int_{\mathbb{R}^n} f(x) \exp(-iw^T x) dx$$

of f , since $\hat{f}(w)$ is the component of f associated with the frequency w .

The key to estimating $\hat{f}(w)$ is to express the generalized transform in terms of the convolution of f with the density function. Since

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

where

$$\rho_\lambda(v) = \frac{1}{\lambda^n} \rho\left(\frac{v}{\lambda}\right),$$

we can write $\langle\langle f \rangle\rangle_\lambda$ as the convolution

$$\langle\langle f \rangle\rangle_\lambda(x) = (f * \rho_\lambda)(x)$$

of f with ρ_λ . If f is integrable, then the Fourier transform of the convolution of f with any integrable function is the product of the convolutions (see, for example, Folland [7, Chapter 8]), thus

$$\widehat{\langle\langle f \rangle\rangle_\lambda}(w) = \widehat{(f * \rho_\lambda)}(w) = \widehat{f}(w)\widehat{\rho_\lambda}(w). \quad (4.1)$$

The following result expresses (4.1) in terms of the original density function ρ .

Theorem 4.1 *If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is integrable, then*

$$\widehat{\langle\langle f \rangle\rangle_\lambda}(w) = \widehat{f}(w)\widehat{\rho}(\lambda w). \quad (4.2)$$

Proof. Since $\langle\langle f \rangle\rangle_\lambda$ is integrable whenever f is integrable, the result follows from (4.1) by verifying that $\widehat{\rho_\lambda}(w) = \widehat{\rho}(\lambda w)$. ■

We use Theorem 4.1 to estimate the rate of decay of the Fourier transform for various density functions. The estimates are simplified when the density function ρ satisfies

$$\rho(x) = \prod_{j=1}^n \rho_j(x_j),$$

where each $\rho_j : \mathbb{R} \mapsto \mathbb{R}$ is a one-dimensional density function, since then we have

$$\widehat{\rho}(w) = \prod_{j=1}^n \widehat{\rho_j}(w_j).$$

This result shows that we can compute $\widehat{\rho}$ once we know the one-dimensional $\widehat{\rho_j}$. The Fourier transform of the one-dimensional uniform density function is obtained directly from the definition; the Fourier transform of the one-dimensional Gaussian density function is a standard calculation. In this manner we obtain that the Fourier transform of the Gaussian density function (2.1) is

$$\widehat{\rho}(w) = \exp\left(-\frac{1}{4}\|w\|^2\right) \quad (4.3)$$

while for the uniform density function (2.7),

$$\widehat{\rho}(w) = \prod_{j=1}^n \frac{\sin w_j}{w_j}. \quad (4.4)$$

Note that $|\widehat{\rho}(w)| \leq 1$ for both density functions, as must be true for any density function.

We use Theorem 4.1 to estimate the rate of decay of the Fourier transform in terms of the largest frequency component. If $|w_{\max}|$ is the largest component of w , then $\|w\| \geq |w_{\max}|$, and thus (4.3) shows that

$$\left| \widehat{\langle\langle f \rangle\rangle_\lambda}(\omega) \right| \leq \exp\left(-\frac{1}{4}\lambda^2|w_{\max}|^2\right) \left| \widehat{f}(\omega) \right|$$

for the Gaussian density function. In contrast, for the uniform density function, (4.4) shows that

$$\left| \langle \widehat{f} \rangle_{\lambda}(\omega) \right| \leq \frac{\sin(\lambda |w_{\max}|)}{\lambda |w_{\max}|} \left| \widehat{f}(\omega) \right|.$$

Clearly, the estimate for the Gaussian density function is more favorable.

Theorem 4.1 and the analysis of the rate of decay of the Fourier transform are an extension of the results obtained by Wu [31] for the Gaussian transform. Although we have considered only the Gaussian and uniform density functions, it is clear that this analysis can be carried out provided we are able to estimate the Fourier transform of the density function.

Figures 4.1 and 4.2 illustrate the smoothing transform as applied to the two dimensional ($n = 2$) version of the Griewank function,

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

This function was constructed by Griewank [9] as a global optimization test function. The Gaussian and uniform transforms for the Griewank function are

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

and

$$\langle \widehat{f} \rangle_{\lambda}(x) = 1 + \sum_{i=1}^n \left(\frac{x_i^2}{200} + \frac{\lambda^2}{600} \right) - \prod_{i=1}^n \left(\sin \left(\frac{\lambda}{\sqrt{i}} \right) / \left(\frac{\lambda}{\sqrt{i}} \right) \right) \cos \left(\frac{x_i}{\sqrt{i}} \right),$$

respectively.

Figures 4.1 and 4.2 show the transformed functions for the Griewank function using the Gaussian and uniform transforms, respectively. The graphs in the left column are the transformed functions, and those in the right are the corresponding contours. The values of λ increase as we go from the graph on the top to the graph in the bottom. For these graphs the values of λ are 0, 1, 2, and 4, with $\lambda = 0$ for the graph on the top. From these pictures, we see that both transforms smooth the function well for increasing λ values. However, for a given λ value, the Gaussian transform is slightly smoother than the uniform transform. In other words, in order to obtain a sufficiently smooth function, a larger λ value may be required for the uniform transform than for the Gaussian transform.

We have applied a simple **Matlab** minimization procedure to the transformed functions in the figures, first the top one, and then the next, and so on. The symbol * in the contours marks the solutions obtained. The pictures show that after smoothing the function, by either the Gaussian or the uniform transform, the global minimizer of the function was found with a few continuation steps.

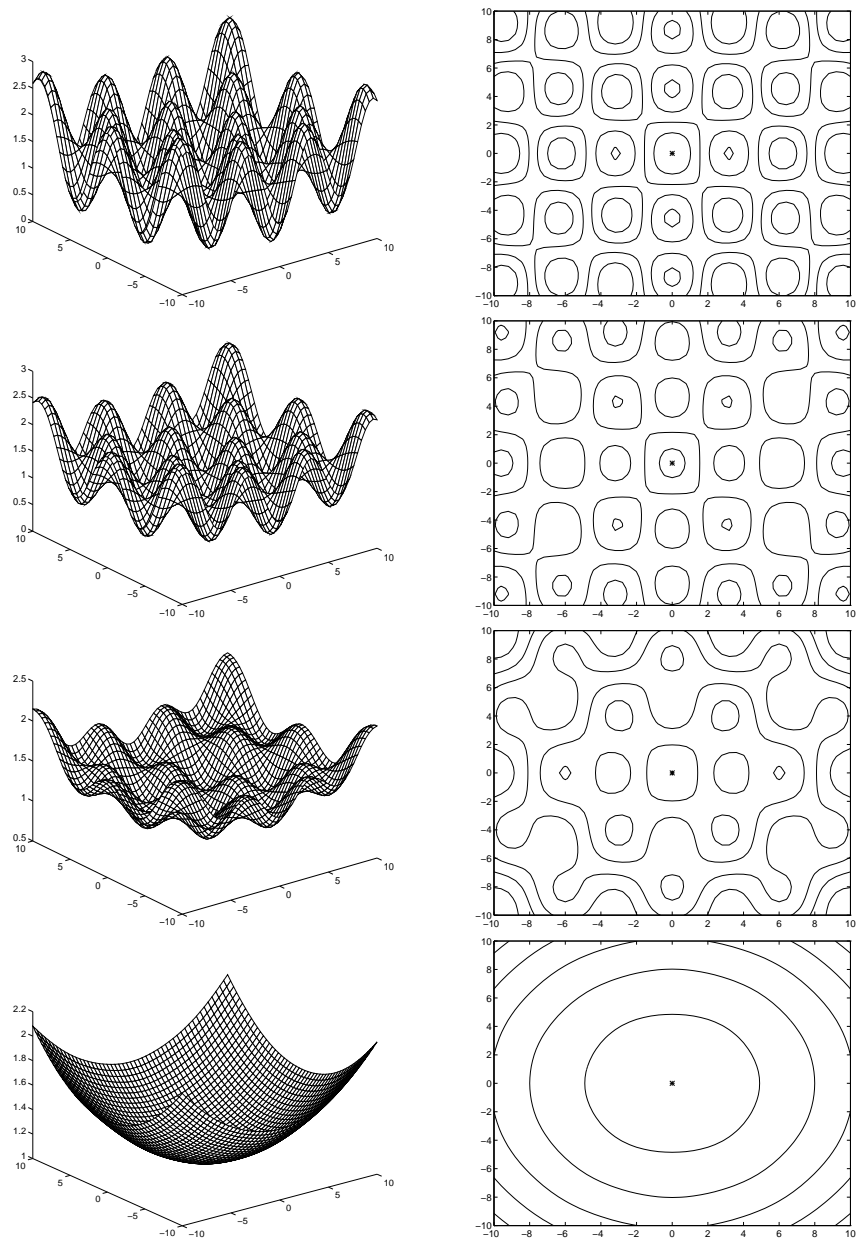


Figure 4.1: The Gaussian transform for the Griewank function

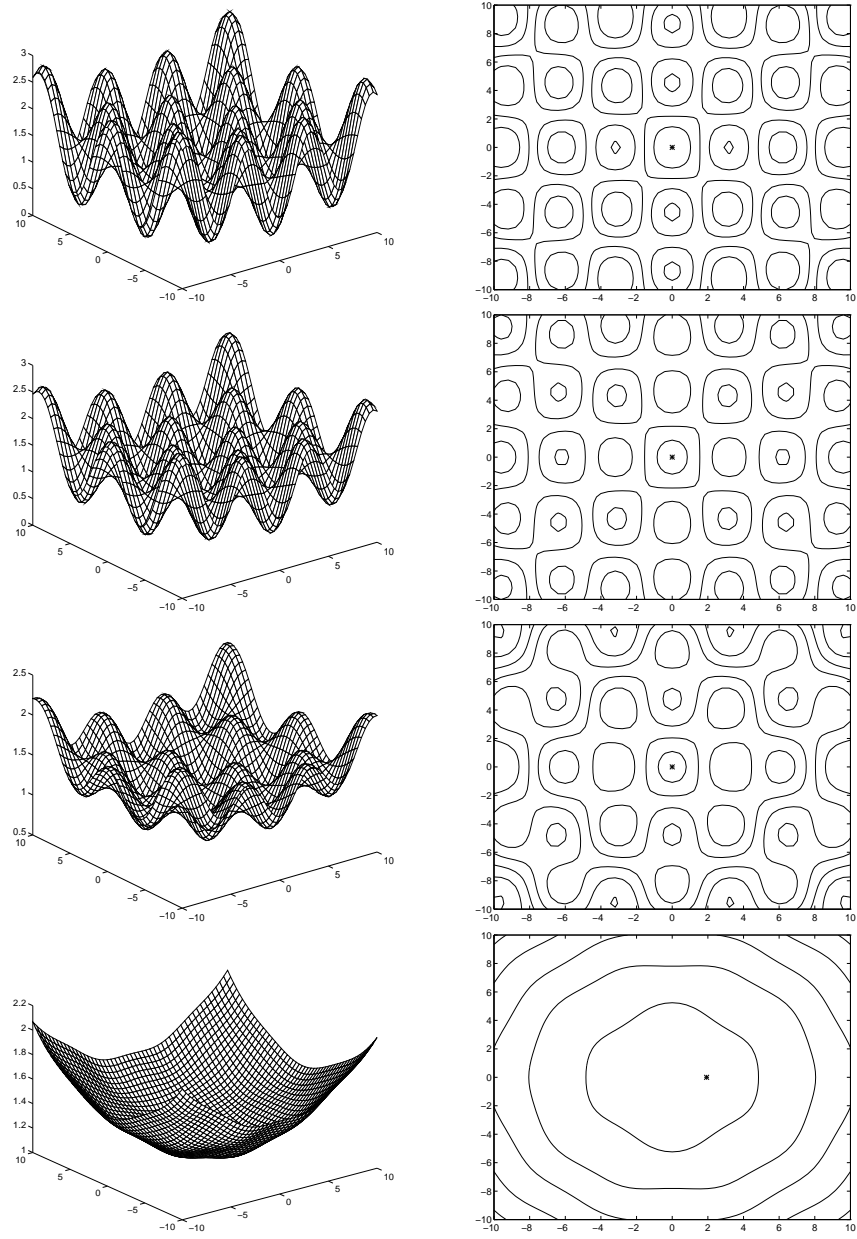


Figure 4.2: The uniform transform for the Griewank function

5 Computing the Generalized Transform

A serious drawback to the general use of the generalized transform is that computing $\langle\langle f \rangle\rangle_\lambda$ for a function defined on \mathbb{R}^n requires the computation of n -dimensional integrals. For functions that arise in applications we can usually avoid this difficulty by reducing the computation of the n -dimensional integrals to k -dimensional integrals where $k \ll n$. In this section we consider a class of functions for which the computation of the generalized transform requires the computation of only 1-dimensional generalized transforms.

Definition 5.1 *A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is decomposable if f can be written in the form*

$$f(x) = \sum_{k=1}^m \left(\prod_{j=1}^n f_{k,j}(x_j) \right), \quad (5.1)$$

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 [31] under the term generalized multilinear functions; Moré and Wu [18] introduced the term decomposable to avoid confusion with the use of multilinear for a function that is linear in each argument.

Decomposable functions form an algebra, that is, a vector space that is closed under multiplication. Linear and quadratic functions in \mathbb{R}^n are decomposable. 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 $p_i \geq 0$, are also decomposable. An interesting example of decomposable functions arises in the fitting of data by minimizing the l_2 norm

$$\Phi(\alpha, x) = \sum_{i=1}^m \left(\sum_{j=1}^n \alpha_j \exp(-t_i x_j) - y_i \right)^2$$

of the difference between an exponential model and the data y_1, \dots, y_m at times t_1, \dots, t_m . Clearly, Φ is a decomposable function of the variables $(\alpha_1, x_1), \dots, (\alpha_n, x_n)$.

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 a one-dimensional Gaussian transform. Indeed, if f is defined by (5.1), 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 $\langle f_{k,j} \rangle_\lambda$.

We extend the above result to the generalized transform $\langle\langle f \rangle\rangle_\lambda$ by imposing an additional assumption on the density function ρ , that is, we assume that

$$\rho(x) = \prod_{j=1}^n \rho_j(x_j), \quad (5.2)$$

where each $\rho_j : \mathbb{R} \mapsto \mathbb{R}$ is a one-dimensional density function. This assumption on ρ means that the random variables associated with each density functions ρ_j are independent. The Gaussian (2.1) and uniform (2.7) density functions, as well as most other n -dimensional density functions, satisfy this assumption.

Theorem 5.2 *If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is decomposable and the density function $\rho : \mathbb{R}^n \mapsto \mathbb{R}$ satisfies (5.2), then*

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

Proof. The result follows from (5.2) and the definition of the generalized transform. ■

Theorem 5.2 can be used to compute the generalized transform of a decomposable function provided we are able to compute the generalized transform of the component functions $f_{k,j}$. In the remainder of this section we explore the case where the component function is analytic.

We obtain an expression for the generalized transform of an analytic function f by noting that in this case the Taylor series converges uniformly, and thus

$$\langle\langle f \rangle\rangle_\lambda(x) = \sum_{l=0}^{+\infty} f^{(l)}(x) \frac{(-1)^j \lambda^j}{j!} \int_{-\infty}^{+\infty} u^l \rho(u) du.$$

This expression simplifies when the density function is even.

Theorem 5.3 *If $f : \mathbb{R} \mapsto \mathbb{R}$ is analytic and $\rho : \mathbb{R} \mapsto \mathbb{R}$ is even, then the generalized transform $\langle\langle f \rangle\rangle_\lambda$ satisfies*

$$\langle\langle f \rangle\rangle_\lambda(x) = \sum_{l=0}^{+\infty} f^{(2l)}(x) \frac{\lambda^{2l}}{(2l)!} \int_{-\infty}^{+\infty} u^{2l} \rho(u) du.$$

Proof. Just note that the integrals with odd powers vanish by symmetry for any density function that is symmetric around the origin. ■

Theorem 5.3 shows that if ρ is the uniform density function then

$$\langle\langle f \rangle\rangle_\lambda(x) = \sum_{l=0}^{+\infty} f^{(2l)}(x) \frac{\lambda^{(2l)}}{(2l+1)!} \quad (5.3)$$

is the generalized transform of f . In particular, for the \sin , \cos , and \exp functions, expression (5.3) shows that

$$\sin(x) \frac{\sin \lambda}{\lambda}, \quad \cos(x) \frac{\sin \lambda}{\lambda}, \quad \exp(x) \frac{\sinh \lambda}{\lambda}$$

are, respectively, the generalized transforms for these functions. Theorem 5.3 also yields an expression for the Gaussian transform if we make use of the identity

$$\frac{1}{\pi^{1/2}} \int_{-\infty}^{+\infty} u^{2l} \exp(-u^2) du = \frac{(2l)!}{4^l l!},$$

which can be verified from standard tables or by integration by parts. Hence, Theorem 5.3 shows that

$$\langle f \rangle_\lambda(x) = \sum_{l=0}^{+\infty} \frac{1}{l!} f^{(2l)}(x) \left(\frac{\lambda}{2}\right)^{(2l)} \quad (5.4)$$

is the Gaussian transform of an analytic function f . As a special case of (5.4),

$$\sin(x) \exp(-\tfrac{1}{4}\lambda^2), \quad \cos(x) \exp(-\tfrac{1}{4}\lambda^2), \quad \exp(x) \exp(\tfrac{1}{4}\lambda^2)$$

are, respectively, the generalized transforms for the \sin , \cos , and \exp functions.

Piela, Kostrowski, and Scheraga [20] used (5.4) as the motivation for the diffusion equation method by noting that the mapping

$$h(x, t) = \langle f \rangle_{2\sqrt{t}}(x) = \sum_{l=0}^{+\infty} \frac{1}{l!} f^{(2l)}(x) t^l$$

satisfies the diffusion equation

$$\frac{\partial^2 h}{\partial x^2}(x, t) = \frac{\partial h}{\partial t}(x, t), \quad h(x, 0) = f(x).$$

Thus, in this approach, the transformation is defined as any solution to the n -dimensional diffusion equation. In later work ([13, 14]) it was shown that (2.4) could also be used to define this transformation in \mathbb{R}^n . In our work we have used (2.4) as the definition of the Gaussian transform and derived all results from this definition.

Theorem 5.4 *If $f : \mathbb{R} \mapsto \mathbb{R}$ is the monic polynomial $f(x) = x^k$, then*

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

is the Gaussian transform, while if ρ is the uniform density function, then

$$\langle\langle f \rangle\rangle_\lambda(x) = \sum_{l=0}^{\lfloor k/2 \rfloor} \left(\frac{k!}{(k-2l)!} \right) \frac{\lambda^{2l}}{(2l+1)!} x^{k-2l}$$

is the generalized transform.

Proof. The result follows directly from (5.3) and (5.4). ■

As an application of Theorem 5.4 to an n -dimensional function, consider the general quadratic

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

where $Q \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}^n$. A computation shows that

$$\langle\langle f \rangle\rangle_\lambda(x) = \frac{1}{2}x^T Q x + c^T x + \frac{\lambda^2}{4} \left(\sum_{i=1}^n q_{i,i} \right)$$

for the Gaussian density function, while

$$\langle\langle f \rangle\rangle_\lambda(x) = \frac{1}{2}x^T Q x + c^T x + \frac{\lambda^2}{6} \left(\sum_{i=1}^n q_{i,i} \right)$$

for the uniform density function. Thus, for quadratic functions, both transforms only differ by a re-scaling of the parameter λ . This result holds for any even density function, since (5.3) shows that

$$\langle\langle f \rangle\rangle_\lambda(x) = x^2 + \lambda^2 \int_{-\infty}^{+\infty} u^2 \rho(u) du$$

when $f(x) = x^2$.

6 The Gaussian Transform for Macromolecular Problems

For macromolecular modeling problems we are interested in transforming a class of functions in terms of the distances between pairs of atoms, where $x_i \in \mathbb{R}^p$ is the position of the i -th atom. In general we are concerned with three-dimensional problems where $p = 3$, but values of $p > 3$ are also of interest. Given functions $f_{i,j} : \mathbb{R}^p \mapsto \mathbb{R}$ of the distance between atoms x_i and x_j , we outline several techniques that can be used to compute the Gaussian transform of the potential function

$$f(x) = \sum_{i,j \in \mathcal{S}} f_{i,j}(x_i - x_j), \quad (6.1)$$

where \mathcal{S} is some subset of all pairs of atoms, and the mappings $f_{i,j}$ are of the form

$$f_{i,j}(x) = h_{i,j}(\|x\|), \quad (6.2)$$

for some mapping $h_{i,j} : \mathbb{R} \mapsto \mathbb{R}$.

The example $x \mapsto \|x\|$ points out that functions of the general form (6.1) and (6.2) are not usually decomposable. On the other hand, we now show that we can still reduce the computation of the Gaussian transform to the computation of one-dimensional integrals in terms of $h_{i,j}$. Note that in these problems f is defined on \mathbb{R}^n , where $n = mp$ and m is the number of atoms, but that $h_{i,j}$ is defined in \mathbb{R} .

The following result of Moré and Wu [19] is needed to prove that computing the Gaussian transform of (6.1) requires only the Gaussian transform of $f_{i,j}$.

Theorem 6.1 *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

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

As an application of Theorem 6.1 consider the mapping $f : \mathbb{R}^n \mapsto \mathbb{R}$ defined by (6.1). Computing the Gaussian transform of this mapping is immediate if we are able to compute the Gaussian transform of $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ defined by

$$f_0(x) = f_{i,j}(x_i - x_j).$$

Clearly, we can find a matrix $P \in \mathbb{R}^{n \times p}$ of the form

$$P = (e_{i_1} - e_{j_1}, \dots, e_{i_p} - e_{j_p})$$

such that $P^T x = x_i - x_j$. Since $P^T P = \sigma^2 I$, where $\sigma^2 = 2$, Theorem 6.1 shows that

$$\langle f_0 \rangle_\lambda(x) = \langle f_{i,j} \rangle_{\sqrt{2}\lambda}(x_i - x_j).$$

An immediate consequence of this result is that

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

is the Gaussian transform of the potential function defined by (6.1). In this case f is defined on \mathbb{R}^n , but $f_{i,j}$ is defined on \mathbb{R}^p .

The other ingredient needed for computing the Gaussian transform of functions defined by (6.1) and (6.2) is the Gaussian transform of the function $f_{i,j}$ defined by (6.2). Note that, unlike Theorem 6.1, the following result requires that f be defined on \mathbb{R}^3 .

Theorem 6.2 *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*

$$\langle f \rangle_\lambda(x) = \frac{1}{\lambda\sqrt{\pi}r} \int_0^\infty sh(s) \left[\exp\left(-\frac{(r-s)^2}{\lambda^2}\right) - \exp\left(-\frac{(r+s)^2}{\lambda^2}\right) \right] ds, \quad (6.3)$$

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

$$\langle f \rangle_\lambda(x) = \frac{1}{\lambda\sqrt{\pi}r} \int_{-\infty}^{+\infty} sh(s) \exp\left(-\frac{(r-s)^2}{\lambda^2}\right) ds. \quad (6.4)$$

Moré and Wu [19] proved Theorem 6.2 and used (6.4) to approximate the Gaussian transform of (1.5). More generally, note that the integrals that appear in (6.3) and (6.4) can be expressed explicitly or in terms of special functions. For example, if $h(s) = s$, then (6.3) can be expressed in terms of the complementary error function

$$x \mapsto \frac{2}{\sqrt{\pi}} \int_x^{+\infty} \exp(-s^2) ds.$$

However, in most cases we need to approximate the integrals with a quadrature. An adaptive quadrature could be used, but this is quite likely to be expensive in terms of function evaluations. Moré and Wu [19] recommended the use of Gaussian quadratures because this leads to a discrete transformation that shares many of the properties of the standard Gaussian transform. For the one-dimensional Gauss transform

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

the use of Gaussian quadratures on this integral yields an approximation

$$\langle f \rangle_{\lambda,q} = \frac{1}{\sqrt{\pi}} \sum_{i=1}^q w_i f(x + \lambda s_i), \tag{6.5}$$

which is exact for all polynomials of degree less than $2q$, where q is the number of nodes in the quadrature. The weights w_i and nodes s_i in (6.5) are independent of f and can be found in the literature (for example, Stroud and Secrest [30]) or can be computed with some of the subroutines in **ORTHOPOL** (Gautschi [8]). 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 Gauss transform. For additional information on Gaussian quadrature, see Stroud and Secrest [30] and Davis and Rabinowitz [5]. Steen, Byrne, and Gelbard [28] have Gaussian quadratures for the integrals in (6.3).

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.

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] P. J. DAVIS AND P. RABINOWITZ, *Methods of Numerical Integration*, Academic Press, second ed., 1984.
- [6] Y. M. ERMOLIEV, V. I. NORKIN, AND R. J.-B. WETS, *The minimization of discontinuous functions: Mollifier subgradients*, SIAM J. Control Optim., 33 (1995), pp. 149–167.
- [7] G. B. FOLLAND, *Real Analysis*, John Wiley & Sons, 1984.
- [8] 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.
- [9] A. GRIEWANK, *Generalized descent for global optimization*, J. Optim. Theory Appl., 34 (1981), pp. 11–39.
- [10] R. W. HASSE AND J. P. SCHIFFER, *The structure of the cylindrically confined coulomb lattice*, Ann. Physics, 203 (1990), pp. 419–448.
- [11] 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.
- [12] V. Y. KATKOVNIK AND O. Y. KULCHITSKII, *Convergence of a class of random search algorithms*, Automat. Remote Control, 8 (1972), pp. 81–87.
- [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. KREIMER AND R. Y. RUBINSTEIN, *Nondifferentiable optimization via smooth approximation: General analytical approach*, Math. Oper. Res., 39 (1992), pp. 97–119.
- [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 Z. WU, *Global continuation for distance geometry problems*, Preprint MCS-P505-0395, Argonne National Laboratory, Argonne, Illinois, 1995a.
- [19] ———, *ε -optimal solutions to distance geometry problems via global continuation*, Preprint MCS-P520-0595, Argonne National Laboratory, Argonne, Illinois, 1995b.
- [20] 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.
- [21] 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.
- [22] R. Y. RUBINSTEIN, *Simulation and the Monte Carlo Method*, John Wiley & Sons, 1981.
- [23] ———, *Smoothed functionals in stochastic optimization*, Math. Oper. Res., 8 (1983), pp. 26–33.
- [24] ———, *Monte Carlo Optimization, Simulation and Sensitivity of Queueing Networks*, John Wiley & Sons, 1986.
- [25] J. P. SCHIFFER, *Phase transitions in anisotropically confined ionic crystals*, Phys. Rev. Lett., 70 (1993), pp. 818–821.
- [26] D. SHALLOWAY, *Application of the renormalization group to deterministic global minimization of molecular conformation energy functions*, J. Global Optim., 2 (1992), pp. 281–311.

- [27] 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.
- [28] 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.
- [29] J. E. STRAUB, *Optimization techniques with applications to proteins*, preprint, Boston University, Department of Chemistry, Boston, Massachusetts, 1994.
- [30] A. H. STROUD AND D. SECREST, *Gaussian Quadrature Formulas*, Prentice-Hall, Inc., 1966.
- [31] 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.