The Effective Energy Transformation Scheme as a General Continuation Approach to Global Optimization with Application to Molecular Conformation

Zhijun Wu[†]

Abstract. This paper discusses a generalization of the special function transformation scheme for global energy minimization for molecular conformation used in [3, 4, 14, 15]. Theories for the method as a general continuation approach are established. We show that the method can transform a nonlinear objective function into a class of gradually deformed, but "smoother", functions. An optimization procedure can then be applied to the new functions successively, to trace the solution back to the original function. Two types of transformation are defined: the isotropic and the anisotropic. We show that both, although not applicable numerically to arbitrary functions because of the required high dimensional integration, can be applied to a large class of nonlinear partially separable functions, and, in particular, the energy functions for molecular conformation. Methods to compute exactly the required transformations are given. Advantages of this transformation approach over the conventional homotopy methods also are discussed.

Abbreviated title: A Continuation Approach to Global Optimization

Key words: global/local minimization, continuation methods, integral transformation, molecular conformation

AMS (MOS) subject classification: 49M37, 68Q22, 92C40

[†]Advanced Computing Research Institute, Cornell University, Ithaca, NY 14853.

1 Introduction

We are interested in solving the global minimization problem associated with the study of molecular conformation, especially protein folding. That is, given a sequence of amino acids, predict the corresponding native, folded, or the most stable, protein structure in the 3-dimensional space.

How protein folds is one of the key biophysical problems of the decade. It is fundamental for almost all theoretical studies of proteins, and also has many applications in the biotechnology industry such as the structure-based drug design for the treatment of polio, AIDS, cancer and other related diseases.

The usual computational approach to the protein folding problem is based on the hypothesis that the protein native structure corresponds to the global minimum of the protein energy. The problem can be attacked computationally by minimizing the protein energy over all possible protein structures, and finding the one with the lowest energy, which is supposed to be the most stable structure of protein.

Mathematically, for a protein molecule of n atoms, let $x = \{x_i \in \mathbf{R}^3,\ i = 1, \ldots, n\}$ represent the molecular structure with each x_i specifying the spatial position of atom i. Then the energy minimization problem for protein folding is to minimize globally a nonlinear function $f(x)$ for all x, i.e.,

$$\mathbf{min}_{x \in \mathbf{S}} \ \ f(x) \tag{1}$$

where $\mathbf{S}$ is the set of all possible molecular structures. The objective function $f(x)$ is the energy function for the given protein molecule, which usually is defined in the following form:

$$f(x) \ = \sum_{i=1, j>i}^{n} h_{ij}(\|r_{ij}\|) \tag{2}$$

where $r_{ij} = x_i - x_j$, and h_{ij} is the pairwise energy function determined by $\|r_{ij}\|$, the distance between atoms i and j. A widely used pairwise potential function is the Van der Waals potential function,

$$h_{ij}(\|r_{ij}\|) = \epsilon_{ij} \left[(\frac{\sigma_{ij}}{\|r_{ij}\|})^{12} - 2(\frac{\sigma_{ij}}{\|r_{ij}\|})^{6} \right] + \frac{q_i q_j}{\epsilon \|r_{ij}\|}, \tag{3}$$

where ϵ_{ij}, σ_{ij}, ϵ, q_i and q_j are all physical constants.

It turns out that Problem (1) is extremely difficult to solve in general. The reason is the following: First, in theory, even simple versions of the problem have been proved to be $\mathcal{NP}$-complete [8]. Second, in practice, the objective function for this problem often contains many, usually exponentially many, local minimizers: therefore search for the global minimizer can be computationally intractable. Finally, the protein molecules tend to be very large, typically containing $\mathcal{O}(10{,}000)$ atoms. For such large problems, if a general global optimization method is used, the required computation will not be affordable even on today's most powerful computers. So the final solution to this problem will definitely require a combination of the most efficient algorithm, the most powerful computer, e.g., parallel computation, and the most effective use of heuristics specific to protein folding.

In response to the challenge of solving the protein folding problem, or the molecular conformation problem in general, many approaches have recently been proposed and studied, either deterministic or undeterministic, such as those in [2, 3, 4, 5, 9, 10, 11, 12, 13, 14, 15, 16, 17]. However, these approaches still can only be applied to small problems, not to the large ones of biological interest. So we believe that developing more efficient and effective algorithms currently is still the most important work to do. The purpose of this work is just to make further progress in this direction.

In this paper we discuss a generalization of the effective energy transformation scheme used in [3, 4, 14, 15] for the global minimization for molecular conformation. Instead of applying the transformation only to the probability distributions of given functions, we now transform the functions directly, generalizing the scheme in [3, 4, 14, 15] to a broader class of functions. Theories for the transformation as a general continuation approach are established. We show that the method can transform a nonlinear objective function into a class of gradually deformed, but "smoother", functions. An optimization procedure can then be applied to the new functions successively, to trace the solution back to the original function. Two types of transformation are defined: the isotropic and the anisotropic. We show that both, although not applicable numerically to arbitrary functions because of the required high dimensional integration, can be applied to a large class of nonlinear partially separable functions, and especially to the energy functions for molecular conformation. The methods to compute exactly the transformation for these functions are given. Advantages of this transformation approach over the conventional homotopy methods also are discussed.

The paper is organized as follows: Section 1 introduces the general approach, and describes the function transformation method. Section 2 studies the mathematical properties of the transformation as a continuation process. Section 3 characterizes the "smoothness" property and shows that the transformed function becomes "smoother" in the sense that the small high frequency variations in the original function are averaged out after the transformation. The numerical applicability of the transformation is discussed in Section 4. The transformation is extended to the anisotropic type in Section 5. The formulas also are derived for computing the transformation exactly for the molecular conformation energy functions. Finally, Section 6 contains concluding remarks.

2 The Approach

In this section, we describe our general approach to global optimization as well as the function transformation scheme to be used.

Suppose that we have a "poorly-behaved" nonlinear function. Because of the "nonsmoothness", this type of functions is very hard to optimize either locally or globally. To overcome the difficulty, we suggest using a special technique to transform the function into a class of gradually deformed, but "smoother", functions like those shown in Figure 1 (a) and (b). The optimization procedure can then be applied to these new functions successively, to trace the solution back to the original function.

To deform the function, we define a function transformation using a special parametrized integral transformation as follows:

Definition 1 *Given a nonlinear function f, the transformation $<f>_\lambda$ for f is defined such that for all x,*

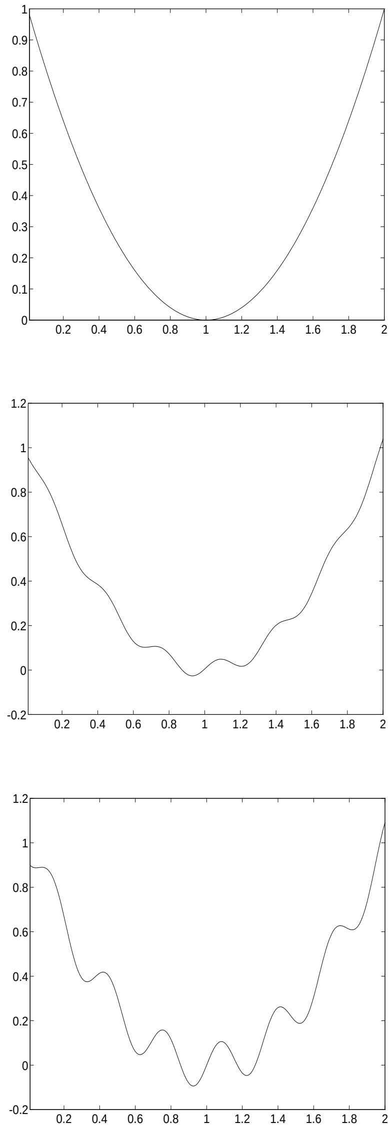
$$<f>_\lambda (x) \;=\; \mathcal{C}_\lambda \int f(x')\; e^{-\|x-x'\|^2/\lambda^2}\; dx', \tag{4}$$

or equivalently,

$$<f>_\lambda (x) \;=\; \mathcal{C}_\lambda \int f(x-x')\; e^{-\|x'\|^2/\lambda^2}\; dx', \tag{5}$$

where λ is a positive number and $\mathcal{C}_\lambda$ is a normalization constant such that

$$\mathcal{C}_\lambda \int e^{-\|x\|^2/\lambda^2}\; dx = 1. \tag{6}$$



(a) $\lambda = 0.2$

(b) $\lambda = 0.1$

(c) $\lambda = 0.0$

Figure 1: A class of gradually deformed functions.

To understand this definition, consider that, given a random function $g(x')$ and a probability distribution function $p(x')$ for the random variable x', the expectation for the function g with respect to p is

$$< g >_p = \int g(x')\ p(x')\ dx'. \tag{7}$$

Compared with this formula, the defined transformation is actually such that the function value for $< f >_\lambda$ at any x is equal to the expectation for f sampled by a Gaussian distribution function centered at x.

For example, consider the following nonlinear function:

$$f(x) = (x-1)^2 + 0.1 sin 20(x-1) \tag{8}$$

which is a quadratic function augmented with some additional noise. The transformation for this function can be computed analytically as follows:

$$< f >_\lambda (x) = (x-1)^2 + \frac{\lambda^2}{2} + 0.1 e^{-(20\lambda)^2/4} sin 20(x-1). \tag{9}$$

The function value $< f >_\lambda (x)$ for fixed x is equal to the integration with respect to two functions, the original function $f(x')$ and the Gaussian distribution function $p(x') = \mathcal{C}_\lambda e^{-\|x-x'\|^2/\lambda^2}$ (see Figure 2), where λ determines the size of the dominant region of the Gaussian. Since the most significant part of the integration is that within the dominant region of the Gaussian, $< f >_\lambda (x)$ can be viewed as the average value for the original function f within a small λ-neighborhood around x. If λ is equal to 0, nothing is averaged and hence the transformed function is exactly the original function. Otherwise, original function variations in small regions are averaged out, and the transformed function will become "smoother".

Figure 1 also shows how the function $< f >_\lambda$ in (9) behaves with increasing λ. As we can see there, when $\lambda = 0.0$, the function is exactly the original function; when we increase λ to 0.1, the function becomes "smoother"; when λ is increased further to 0.2, the function becomes entirely "smooth". As we will show in the following sections, what we observed here is actually a general property for the transformation, i.e., for any function f, the larger the λ value is set to, the "smoother" the transformed function will be.

Figure 3 illustrates what the transformation we introduced above could imply for optimization. We applied a standard optimization procedure, the

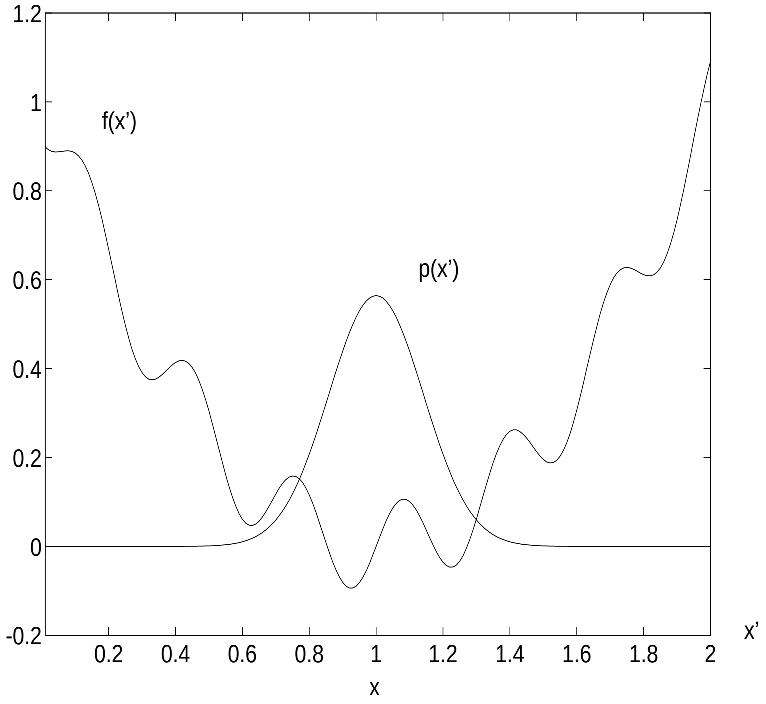


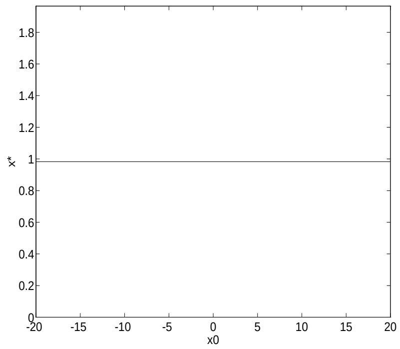
Figure 2: A 1-dimensional transformation example.

quasi-Newton's method with line search, to the three functions in Figure 1, respectively. Figure 3 (a), (b), and (c) contain the corresponding solutions x^* we obtained with different choices of initial guess x°. Although globally convergent, the method can not always find the right solution if the "noise" is large. So for the function in Figure 1 (c), the method converged to the right solution only when the initial guess was close enough to the solution. With the initial guess far from the solution, the method failed to find the right solution (Figure 3 (c)). For the function in Figure 1 (b), although it is "smoother", the method still behaved almost the same as before (Figure 3 (b)). But for the function in Figure 1 (a), the method always converged to the right solution (Figure 3 (a)), as we can expect. However, if we apply the procedure to the functions in Figure 1 (a) to (c) successively, and at each step, take the solution for the previous function as the starting point, the solutions for all these functions can then be obtained whatever the initial guess is.

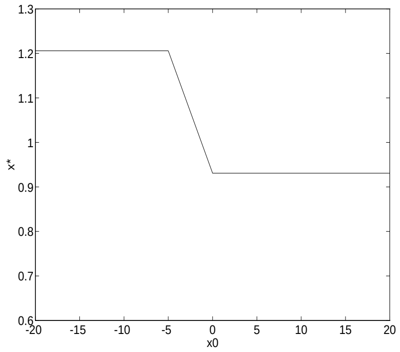
The experiment above suggests a general optimization process. Specifically, to optimize a difficult function, either locally or globally, one can use the transformation technique to deform the function into a class of "smoother" or "easier" functions, and apply the optimization procedure to the functions successively, to trace the solution back to the original function.

3 Continuation

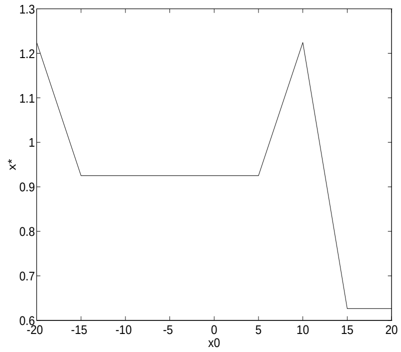
One might wonder what is the difference between the suggested approach and the general homotopy methods. The answer is that this approach is indeed a special type of homotopy method. But the transformation is different from conventional homotopies, and has the following three special features: First, the transformed functions are not arbitrarily deformed functions. They all are approximations to the original function in the sense that they are coarse estimates for the latter. Second, the transformation is defined by a special parametrized integral transformation. If we increase the value of the parameter, the transformed function will become "smoother" with small function variations gradually removed. However the overall structure of the original function will still maintain. Finally, if we apply the optimization procedure to a transformed function, the solution for it should already be very close to the solution for the original function. All these features are



(a)



(b)



(c)

Figure 3: The solutions for the functions in Figure 1 obtained by the quasi-Newton's method with different initial guesses.

good for global optimization (also for robust local optimization), but not necessarily the properties of conventional homotopies.

We show in the following that the proposed transformation is indeed a well defined homotopy, and determines for any initial solution a unique solution curve containing the stationary points for the transformed functions.

Assumption 1 *The objective function f is twice continuously differentiable, and the transformation $<>_\lambda$ is well defined for the function as well as its derivatives.*

Assumption 2 *Let g be the gradient of f, and Δ the Laplace operator,*

$$\Delta = \sum_i \frac{\partial^2}{\partial x_i^2}. \tag{10}$$

Then the operation Δ can be applied to g, and the transformation $<>_\lambda$ is well defined for all involved derivatives. Also, $\Delta g(x)$ is uniformly bounded and satisfies a Lipschitz condition:

$$\|\Delta g(x_1) - \Delta g(x_2)\| \leq L \ \|x_1 - x_2\| \qquad \forall x_1,\ x_2. \tag{11}$$

Assumption 3 *The transformation $< \nabla^2 f >_\lambda (x)$ is uniformly bounded away from 0 and also satisfies a Lipschitz condition:*

$$\| < \nabla^2 f >_\lambda (x_1) - < \nabla^2 f >_\lambda (x_2)\| \leq L \ \|x_1 - x_2\| \qquad \forall x_1,\ x_2. \tag{12}$$

We first state two sets of standard results for the proposed transformation in the following lemmas without proof:

Lemma 1 *Under Assumption 1, $\forall \lambda, x$,*

$$\begin{aligned} \nabla < f >_\lambda (x) &= < \nabla f >_\lambda (x) \qquad (13) \\ \nabla^2 < f >_\lambda (x) &= < \nabla^2 f >_\lambda (x). \qquad (14) \end{aligned}$$

Lemma 2 *Under Assumption 1, $\forall x$,*

$$\begin{aligned} \lim_{\lambda \to 0} < f >_\lambda (x) &= f(x) \qquad (15) \\ \lim_{\lambda \to 0} \nabla < f >_\lambda (x) &= \nabla f(x) \qquad (16) \\ \lim_{\lambda \to 0} \nabla^2 < f >_\lambda (x) &= \nabla^2 f(x). \qquad (17) \end{aligned}$$

For the convenience, we define a function $h(\lambda, x)$, $\lambda \in \Gamma$ and $x \in S$, such that $\forall(\lambda, x) \in \Gamma \times S$,

$$h(\lambda, x) = <f>_\lambda (x), \tag{18}$$

where $\Gamma = [0, \lambda_0]$ for some $\lambda_0 < \infty$ and S is a vector space. With this definition, the condition for x to be a stationary point of $<f>_\lambda$ for some $\lambda \in \Gamma$ becomes

$$h_x^{'}(\lambda, x) = 0. \tag{19}$$

Theorem 1 *Let f be a 1-dimensional function and h be defined as in (18). Then under Assumptions 1 and 2, $h_{x\lambda}^{''}(\lambda, x)$ exists and is uniformly bounded for all $\lambda \in \Gamma$ and $x \in S$, and also satisfies a Lipschitz condition in x:*

$$|h(\lambda, x_1) - h(\lambda, x_2)| \leq L \; |x_1 - x_2| \qquad \forall x_1, x_2 \in S. \tag{20}$$

In addition,

$$h_{x\lambda}^{''}(\lambda, x) = \frac{\lambda}{2} < f''' >_\lambda (x). \tag{21}$$

Proof: Let $p(\lambda, x)$ be the Gaussian distribution function defined as follows:

$$p(\lambda, x) = c_\lambda \; e^{-x^2/\lambda^2} \tag{22}$$

where $c_\lambda = 1/(\lambda\sqrt{\pi})$. Then by the definition of $<f>_\lambda$,

$$h(\lambda, x) = <f>_\lambda (x) = \int f(x')p(\lambda, (x - x'))dx'. \tag{23}$$

By Lemma 1,

$$h_x^{'}(\lambda, x) = \int g(x')p(\lambda, (x - x'))dx'. \tag{24}$$

It follows after differentiating (24) with respect to λ that

$$\begin{aligned} h_{x\lambda}^{''}(\lambda, x) &= \int g(x')(\frac{2(x - x')^2}{\lambda^3} - \frac{1}{\lambda})p(\lambda, (x - x'))dx' \\ &= \frac{1}{\lambda}\int (\frac{2}{\lambda^2}(x - x')^2 - 1)g(x')p(\lambda, (x - x'))dx' \\ &= \frac{1}{\lambda}(\frac{2}{\lambda^2}(\Delta_1 - \Delta_2) - <g>_\lambda (x)) \end{aligned} \tag{25}$$

where

$$\Delta_1 = \int x(x-x')g(x')p(\lambda,(x-x'))dx' \qquad (26)$$
$$\Delta_2 = \int x'(x-x')g(x')p(\lambda,(x-x'))dx'. \qquad (27)$$

It is easy to verify that

$$\Delta_1 = -\frac{\lambda^2}{2}x <g'>_\lambda (x) \qquad (28)$$
$$\Delta_2 = -\frac{\lambda^2}{2}(<g>_\lambda (x)+ <xg'>_\lambda (x)). \qquad (29)$$

Also note that

$$\begin{aligned} x <g'>_\lambda &- <xg'>_\lambda \\ &= \int (x-x')g'(x')p(\lambda,(x-x'))dx' \\ &= \frac{\lambda^2}{2}\int g'(x')dp(\lambda,(x-x')) \\ &= \frac{\lambda^2}{2}(g'(x')p(\lambda,(x-x'))|_{-\infty}^{+\infty} - \int g''(x')p(\lambda,(x-x'))dx') \\ &= -\frac{\lambda^2}{2} <g''> (x). \qquad (30) \end{aligned}$$

Then

$$\begin{aligned} \Delta_1 - \Delta_2 &= -\frac{\lambda^2}{2}x <g'>_\lambda (x) + \frac{\lambda^2}{2}(<g>_\lambda (x)+ <xg'>_\lambda (x)) \\ &= -\frac{\lambda^2}{2}(x <g'>_\lambda - <xg'>_\lambda) + \frac{\lambda^2}{2} <g>_\lambda (x) \\ &= \frac{\lambda^4}{4} <g''>_\lambda (x) + \frac{\lambda^2}{2} <g>_\lambda (x). \qquad (31) \end{aligned}$$

Replacing $(\Delta_1 - \Delta_2)$ in (25) by (31), we see that

$$\begin{aligned} h''_{x\lambda}(\lambda,x) &= \frac{1}{\lambda}(\frac{\lambda^2}{2} <g''>_\lambda (x)+ <g>_\lambda (x)- <g>_\lambda (x)) \\ &= \frac{\lambda}{2} <f'''>_\lambda (x). \qquad (32) \end{aligned}$$

By Assumption 2, $<f'''>_\lambda (x)$ is well defined and uniformly bounded. So, $h''_{x\lambda}(\lambda, x)$ exists and uniformly bounded for all $\lambda \in \Gamma$ and $x \in S$. Since $f'''(x)$ satisfies a Lipschitz condition by Assumption 2, $<f'''>_\lambda (x)$ also satisfies a Lipschitz condition:

$$| <f'''>_\lambda (x_1) - <f'''>_\lambda (x_2)| \leq L_0 \; |x_1 - x_2| \qquad \forall x_1, \; x_2. \tag{33}$$

Let $L = \frac{\lambda_0}{2} L_0$. Then (20) follows immediately. So $h''_{x\lambda}(\lambda, x)$ satisfies a Lipschitz condition in x. $\square$

Theorem 2 *Let f be an n-dimensional function and h be defined as in (18). Then under Assumptions 1 and 2, $h''_{x\lambda}(\lambda, x)$ exists and is uniformly bounded for all $\lambda \in \Gamma$ and $x \in S$, and also satisfies a Lipschitz condition in x:*

$$\|h''_{x\lambda}(\lambda, x_1) - h''_{x\lambda}(\lambda, x_2)\| \leq L \; \|x_1 - x_2\| \qquad \forall x_1, x_2 \in S. \tag{34}$$

In addition,

$$h''_{x\lambda}(\lambda, x) = \frac{\lambda}{2} < \Delta g >_\lambda (x). \tag{35}$$

Proof: Let $p_n(\lambda, x)$ be the Gaussian distribution function

$$p_n(\lambda, x) = c_\lambda^n \; e^{-\|x\|^2/\lambda^2} \tag{36}$$

where $c_\lambda^n = 1/(\lambda\sqrt{\pi})^n$. Then by the definition of $<f>_\lambda$,

$$h(\lambda, x) = <f>_\lambda (x) = \int f(x') p_n(\lambda, (x - x'))dx'. \tag{37}$$

By Lemma 1,

$$h'_x(\lambda, x) = \int g(x') p_n(\lambda, (x - x'))dx'. \tag{38}$$

Differentiate (38) with respect to λ to obtain

$$\begin{aligned} h''_{x\lambda}(\lambda, x) &= \int g(x')(\frac{2\|x - x'\|^2}{\lambda^3} - \frac{n}{\lambda}) p_n(\lambda, (x - x'))dx' \\ &= \frac{1}{\lambda}(\frac{2}{\lambda^2}\sigma - n <g>_\lambda (x)), \end{aligned} \tag{39}$$

where

$$\begin{aligned}\sigma &= \int \|x - x'\|^2 g(x') p_n(\lambda, (x - x')) dx' \\ &= \sum_{i=1}^{n} \int (x_i - x_i')^2 g(x') p_n(\lambda, (x - x')) dx' \\ &= \sum_{i=1}^{n} \sigma_i. \end{aligned} \tag{40}$$

Let $\bar{x}_i = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)^T$. Then

$$\begin{aligned}\sigma_i &= \int (\int (x_i - x_i')^2 g(x_i', \bar{x}_i') p(\lambda, (x_i - x_i')) dx_i') \\ &\quad p_{n-1}(\lambda, (\bar{x}_i - \bar{x}_i')) d\bar{x}_i' \\ &= \int (\Delta_{1i} - \Delta_{2i}) p_{n-1}(\lambda, (\bar{x}_i - \bar{x}_i')) d\bar{x}_i' \end{aligned} \tag{41}$$

where

$$\Delta_{1i} = \int x_i (x_i - x_i') g(x_i', \bar{x}_i') p(\lambda, (x_i - x_i')) dx_i' \tag{42}$$

$$\Delta_{2i} = \int x_i' (x_i - x_i') g(x_i', \bar{x}_i') p(\lambda, (x_i - x_i')) dx_i'. \tag{43}$$

From the proof for Theorem 1,

$$\begin{aligned}\Delta_{1i} - \Delta_{2i} &= \frac{\lambda^4}{4} \int \frac{\partial^2 g}{\partial x_i^2}(x_i', \bar{x}_i') p(\lambda, (x_i - x_i')) dx_i' \\ &+ \frac{\lambda^2}{2} \int g(x_i', \bar{x}_i') p(\lambda, (x_i - x_i')) dx_i'. \end{aligned} \tag{44}$$

Substitute (44) back into (41) to obtain

$$\sigma_i = \frac{\lambda^4}{4} < \frac{\partial^2 g}{\partial x_i^2} >_\lambda (x) + \frac{\lambda^2}{2} < g >_\lambda (x). \tag{45}$$

Then

$$\sigma = \frac{\lambda^4}{4} \sum_{i=1}^{n} < \frac{\partial^2 g}{\partial x_i^2} >_\lambda (x) + \frac{n\lambda^2}{2} < g >_\lambda (x) \tag{46}$$

and

$$\begin{aligned}
h''_{x\lambda}(\lambda, x) &= \frac{\lambda}{2}\sum_{i=1}^{n} < \frac{\partial^2 g}{\partial x_i^2} >_\lambda (x) \\
&= \frac{\lambda}{2} < \sum_{i=1}^{n} \frac{\partial^2 g}{\partial x_i^2} >_\lambda (x) \\
&= \frac{\lambda}{2} < \Delta g >_\lambda (x). \qquad (47)
\end{aligned}$$

Then similar to the proof for Theorem 1, it follows immediately that $h''_{x\lambda}(\lambda, x)$ exists and is uniformly bounded for all $\lambda \in \Gamma$ and $x \in S$, and also satisfies a Lipschitz condition in x. $\square$

Finally we state and prove the main theorem in this section as follows:

Theorem 3 *Let f be a function for which Assumptions 1, 2 and 3 all hold. Then for any stationary point x_0 of $< f >_{\lambda_0}$, there is a continuous and differentiable curve $x(\lambda)$, $\lambda \in \Gamma$, such that $x_0 = x(\lambda_0)$ and $\forall \lambda \in \Gamma$, $x(\lambda)$ is a stationary point of $< f >_\lambda$. The curve $x(\lambda)$ is also the unique solution of the initial value problem*

$$\begin{aligned}
x' &= -\frac{\lambda}{2} < \nabla^2 f >_\lambda^{-1} (x) < \Delta g >_\lambda (x) \qquad (48) \\
x_0 &= x(\lambda_0). \qquad (49)
\end{aligned}$$

Proof: Since x_0 is a stationary point of $< f >_{\lambda_0}$,

$$h'_x(\lambda_0, x_0) = 0. \qquad (50)$$

By Assumptions 1, 2, 3, Lemmas 1, 2, and Theorem 2, function h'_x is continuously differentiable at all $(\lambda, x) \in \Gamma \times S$. So by the Implicit Function Theorem, there is a continuously differentiable function $x(\lambda)$ at a neighborhood of λ_0, such that $x_0 = x(\lambda_0)$ and

$$h'_x(\lambda, x(\lambda)) = 0 \qquad (51)$$

for all λ in the neighborhood.

We now show that $x(\lambda)$ also is defined uniquely in Γ.

By differentiating (51), we see that $x(\lambda)$ is a solution to the initial value problem:

$$\begin{aligned} x' &= -h''_{xx}(\lambda, x)^{-1} h''_{x\lambda}(\lambda, x) && (52) \\ x_0 &= x(\lambda_0) && (53) \end{aligned}$$

which, by Lemma 1 and Theorem 2, is equivalent to the problem (48)-(49). Then it suffices to show that the right-hand side of (52) satisfies a Lipschitz condition in x on $\Gamma \times S$, which guarantees a unique solution $x(\lambda)$ in Γ by the standard ordinary differential equation theory.

Under Assumption 3, for $h''_{xx}(\lambda, x)^{-1}$, $\exists M_1, L_1$, such that

$$\begin{aligned} \|h''_{xx}(\lambda, x)^{-1}\| &\leq M_1 \qquad \forall (\lambda, x) \in \Gamma \times S && (54) \\ \|h''_{xx}(\lambda, x_1)^{-1} - h''_{xx}(\lambda, x_2)^{-1}\| &\leq L_1 \, \|x_1 - x_2\| \qquad \forall x_1, x_2 \in S. && (55) \end{aligned}$$

By Theorem 2, for $h''_{x\lambda}(\lambda, x)$, $\exists M_2, L_2$, such that

$$\begin{aligned} \|h''_{x\lambda}(\lambda, x)\| &\leq M_2 \qquad \forall (\lambda, x) \in \Gamma \times S && (56) \\ \|h''_{x\lambda}(\lambda, x_1) - h''_{x\lambda}(\lambda, x_2)\| &\leq L_2 \, \|x_1 - x_2\| \qquad \forall x_1, x_2 \in S. && (57) \end{aligned}$$

Let

$$G(\lambda, x) = -h''_{xx}(\lambda, x)^{-1} h''_{x\lambda}(\lambda, x). \tag{58}$$

Then it is easy to verify that $G(\lambda, x)$ satisfies a Lipschitz condition in x on $\Gamma \times S$:

$$\|G(\lambda, x_1) - G(\lambda, x_2)\| \leq L \, \|x_1 - x_2\| \qquad \forall x_1, x_2 \in S \tag{59}$$

with $L = M_1 L_2 + M_2 L_1$, which completes the proof. $\square$

4 "Smoothness"

We showed by examples in Section 2 that the transformed functions are "smoother" than the original function in the sense that they vary slower and may even have fewer local minimizers. In the following, we characterize more rigorously the "smoothness" of the transformation.

Let $\hat{f}$ be the Fourier transformation for function f, and $\widehat{<f>_\lambda}$ for function $<f>_\lambda$. Recall that the transformation $<f>_\lambda$ for f is just a convolution of f and p, where p is the Gaussian distribution function

$$p(x) = \mathcal{C}_\lambda \ e^{-\|x\|^2/\lambda^2}. \tag{60}$$

Then the Fourier transformation for $<f>_\lambda$ is equal to the product of the Fourier transformations for f and p. The Fourier transformation for the Gaussian distribution function is

$$\hat{g}(\omega) = e^{-\frac{\lambda^2\|\omega\|^2}{4}}. \tag{61}$$

So, we have

$$\widehat{<f>}_\lambda(\omega) = e^{-\frac{\lambda^2\|\omega\|^2}{4}} \hat{f}(\omega). \tag{62}$$

We see from the equality in (62) that if $\lambda \to 0$, $\widehat{<f>}_\lambda$ converges to $\hat{f}$, and so does $<f>_\lambda$ to f. This is exactly the fact we stated in Lemma 2.

Also by the equality in (62), for fixed λ, if ω is large, the exponential term $e^{-\frac{\lambda^2\|\omega\|^2}{4}}$ and hence $\widehat{<f>}_\lambda(\omega)$ will be very small, which implies that high frequency components of the original function become very small after the transformation. This is why the transformed function is "smoother". In addition, for larger λ values, wider ranges of high frequency components of the original function practically vanish after the transformation, and therefore, the transformed function becomes "smoother" and "smoother" with increasing λ values. We state these properties formally in the following theorem.

Theorem 4 *Let f, $\hat{f}$, $<f>_\lambda$ and $\widehat{<f>}_\lambda$ all be given and well defined. Then $\forall \varepsilon > 0$, $\exists \delta \propto 1/\lambda$ for fixed λ, such that $\forall \omega$, $\|\omega\| > \delta$,*

$$\frac{|\widehat{<f>}_\lambda(\omega)|}{|\hat{f}(\omega)|} < \varepsilon. \tag{63}$$

Proof: Given $\varepsilon > 0$, for fixed λ, let $\delta = 2\sqrt{\ln{(1/\varepsilon)}}/\lambda$. Then $\forall \omega$, $\|\omega\| > \delta$,

$$\frac{|\widehat{<f>}_\lambda(\omega)|}{|\hat{f}(\omega)|} = e^{-\frac{\lambda^2\|\omega\|^2}{4}} < e^{-\frac{\lambda^2\delta^2}{4}} = \varepsilon. \tag{64}$$

$\square$

From Theorem 4 we learn that the relative size of $\widehat{<f>_\lambda}(\omega)$ can be made arbitrarily small for all ω with $\|\omega\|$ greater than a value inversely proportional to λ. So if larger λ values are used, more such ω can be included, i.e., more high frequency components can be removed.

5 Numerical Applicability

Practically, the proposed transformation scheme can not be applied to arbitrary functions because it involves high dimensional integration, which can not be computed numerically in general. However, in this section we show the method can be applied to a large class of nonlinear partially separable functions, and especially to typical molecular conformation energy functions.

We state several useful properties for the transformation in the following: First, for the sum of the functions

$$f = \sum_i f_i, \tag{65}$$

the transformation for f is equal to the sum of the transformations for f_i's:

$$<f>_\lambda = \sum_i <f_i>_\lambda. \tag{66}$$

Second, for the multiple of the functions

$$g = \prod_i g_i, \tag{67}$$

where g_i's do not share common variables, the transformation for g is equal to the multiple of the transformations for g_i's:

$$<g>_\lambda = \prod_i <g_i>_\lambda. \tag{68}$$

Finally, for a large class of nonlinear partially separable functions, called the generalized multilinear functions,

$$f = \sum_i \prod_j g_j^i, \tag{69}$$

where g_j^i's are one dimensional nonlinear functions, we have

$$< f >_\lambda = \sum_i \prod_j < g_j^i >_\lambda . \tag{70}$$

Since transformation $< g_j^i >_\lambda$ for all i and j involves only one dimensional integration which can be done numerically, the transformation for a generalized multilinear function can be computed.

Now let us consider a typical n-atom molecular conformation energy function,

$$f(x) \ = \sum_{i=1, j>i}^{n} h_{ij}(\|x_i - x_j\|) \tag{71}$$

where $x = \{x_i \in \mathbf{R}^3, \ i = 1, \ldots, n\}$ and h_{ij} is the pairwise energy function determined by $\|x_i - x_j\|$, the distance between atoms i and j. By (66), the transformation for this energy function is equal to the sum of the transformations for the pairwise energy functions. However the computation for the pairwise transformation still cannot be conducted directly, because there are still more than one variables. Nevertheless, the following theorem provides a feasible way to compute the molecular energy transformation:

Theorem 5 *Let f be defined as in (71). Then the transformation for f can be computed using the formula*

$$< f >_\lambda (x) = \sum_{i=1, j>i}^{n} < h_{ij} >_{\sqrt{2}\lambda} (\|r_{ij}\|) \tag{72}$$

where $r_{ij} = x_i - x_j$ and

$$< h_{ij} >_{\sqrt{2}\lambda} (\|r_{ij}\|) = c_{\sqrt{2}\lambda} \int h_{ij}(\|r'_{ij}\|) e^{-\|r_{ij} - r'_{ij}\|^2/2\lambda^2} dr'_{ij}. \tag{73}$$

Proof: By the definition of $< f >_\lambda$, for any x,

$$\begin{aligned} & < f >_\lambda (x) \\ = & \sum_{i=1, j>i}^{n} < h_{ij} >_\lambda (\|x_i - x_j\|) \\ = & \sum_{i=1, j>i}^{n} c_\lambda^2 \int h_{ij}(\|x'_i - x'_j\|) e^{-\frac{\|x_i - x'_i\|^2 + \|x_j - x'_j\|^2}{\lambda^2}} dx'_i dx'_j, \end{aligned} \tag{74}$$

where c_λ is such that

$$c_\lambda \int e^{-\|x_i\|^2/\lambda^2} dx_i = 1, \qquad \forall i. \tag{75}$$

Make the following variable transformation:

$$r_{ij} = x_i - x_j, \qquad \bar{r}_{ij} = x_i + x_j. \tag{76}$$

Then it is easy to verify that

$$\begin{aligned}
< f >_\lambda (x) &= \sum_{i=1, j>i}^{n} c_\lambda^2 \int h_{ij}(\|x_i' - x_j'\|) e^{-\frac{\|x_i - x_i'\|^2 + \|x_j - x_j'\|^2}{\lambda^2}} dx_i' dx_j' \\
&= \sum_{i=1, j>i}^{n} c_{\sqrt{2}\lambda} \int h_{ij}(\|r_{ij}'\|) e^{-\|r_{ij} - r_{ij}'\|^2/2\lambda^2} dr_{ij}' \\
&= \sum_{i=1, j>i}^{n} < h_{ij} >_{\sqrt{2}\lambda} (\|r_{ij}\|).
\end{aligned} \tag{77}$$

The integral in $< h_{ij} >_{\sqrt{2}\lambda} (\|r_{ij}\|)$ involves only variable r_{ij} which, in spherical coordinates, can further be reduced to a one dimensional integral (see [14]). The transformation $< f >_\lambda (x)$ can then be computed exactly using (72). □

6 Anisotropic vs. Isotropic

The transformation we discussed so far is of isotropic type in the sense that it averages function variations equally along all directions in the variable space. In practice, we might want to average different sizes of function variations along different directions, i.e., use different λ values for different variables, in order to obtain more accurate overall structure of the function. For this purpose, we can define a more general transformation, called the anisotropic transformation, in the following:

Definition 2 *Given a nonlinear function f, the anisotropic transformation $< f >_\Lambda$ for f is defined such that for all x,*

$$< f >_\Lambda (x) = \mathcal{C}_\Lambda \int f(x') \, e^{-\|\frac{x-x'}{\Lambda}\|^2} \, dx', \tag{78}$$

or equivalently,

$$< f >_\Lambda (x) = \mathcal{C}_\Lambda \int f(x - x') \, e^{-\|\frac{x'}{\Lambda}\|^2} \, dx', \tag{79}$$

where Λ *is a diagonal matrix with positive diagonal elements:*

$$\Lambda = \mathbf{diag}([\lambda_1, \lambda_2 \ldots]), \tag{80}$$

and $\mathcal{C}_\Lambda = \prod_i c_{\lambda_i}$ *with* c_{λ_i} *determined such that*

$$c_{\lambda_i} \int e^{-\|x_i\|^2/\lambda_i^2}\, dx_i = 1, \qquad \forall i. \tag{81}$$

From this definition, we see that the anisotropic transformation will be reduced to the isotropic one when the diagonal elements of Λ are all identical.

It turns out that for this more general transform we still have all properties similar to (66), (68) and (70) for the isotropic transform except for the cases of (68) and (70) where different λ values are used for different element functions. We state these properties in the following:

First, for the sum of the functions

$$f = \sum_i f_i, \tag{82}$$

we have

$$< f >_\Lambda = \sum_i < f_i >_\Lambda. \tag{83}$$

Second, for the multiple of the functions

$$g = \prod_i g_i, \tag{84}$$

where g_i's do not share common variables, we have

$$< g >_\Lambda = \prod_i < g_i >_{\Lambda_i}, \tag{85}$$

where Λ_i's are small diagonal matrices and $\Lambda = \mathbf{diag}([\Lambda_1, \Lambda_2, \ldots])$.

Third, for the generalized multilinear functions,

$$f = \sum_i \prod_j g_j^i, \tag{86}$$

where g_j^i's are one dimensional nonlinear functions, we have

$$< f >_\Lambda = \sum_i \prod_j < g_j^i >_{\lambda_j}. \tag{87}$$

We can also derive a simple formula to compute the anisotropic transformation for the molecular conformation energy function:

Theorem 6 *Let f be defined as in (71). Then the anisotropic transformation for f can be computed using the formula*

$$< f >_\Lambda (x) = \sum_{i=1, j>i}^{n} < h_{ij} >_{\sqrt{\lambda_i^2+\lambda_j^2}} (\|r_{ij}\|) \tag{88}$$

where $r_{ij} = x_i - x_j$ and

$$< h_{ij} >_{\sqrt{\lambda_i^2+\lambda_j^2}} (\|r_{ij}\|) = c_{\sqrt{\lambda_i^2+\lambda_j^2}} \int h_{ij}(\|r'_{ij}\|) e^{-\frac{\|r_{ij}-r'_{ij}\|^2}{\lambda_i^2+\lambda_j^2}} dr'_{ij}. \tag{89}$$

Proof: By the definition of $< f >_\Lambda$, for any x,

$$\begin{aligned} & < f >_\Lambda (x) \\ = & \sum_{i=1, j>i}^{n} < h_{ij} >_\Lambda (\|x_i - x_j\|) \\ = & \sum_{i=1, j>i}^{n} c_{\lambda_i} c_{\lambda_j} \int h_{ij}(\|x'_i - x'_j\|) e^{-(\frac{\|x_i - x'_i\|^2}{\lambda_i^2} + \frac{\|x_j - x'_j\|^2}{\lambda_j^2})} dx'_i dx'_j, \end{aligned} \tag{90}$$

where c_{λ_i} is such that

$$c_{\lambda_i} \int e^{-\|x_i\|^2/\lambda^2} dx_i = 1, \qquad \forall i. \tag{91}$$

Make the following variable transformation:

$$r_{ij} = x_i - x_j, \qquad \bar{r}_{ij} = \lambda_j^2 x_i + \lambda_i^2 x_j. \tag{92}$$

Then we have

$$\frac{x_i}{\lambda_i} = \frac{1}{\lambda_i^2 + \lambda_j^2} (\frac{\bar{r}_{ij}}{\lambda_i} + \lambda_i r_{ij}) \tag{93}$$

$$\frac{x_j}{\lambda_j} = \frac{1}{\lambda_i^2 + \lambda_j^2} (\frac{\bar{r}_{ij}}{\lambda_j} - \lambda_j r_{ij}). \tag{94}$$

Using these relations we can verify that

$$< f >_\Lambda (x) = \sum_{i=1, j>i}^{n} c_{\lambda_i} c_{\lambda_j} \int h_{ij}(\|x'_i - x'_j\|) e^{-(\frac{\|x_i - x'_i\|^2}{\lambda_i^2} + \frac{\|x_j - x'_j\|^2}{\lambda_j^2})} dx'_i dx'_j$$

$$= \sum_{i=1,j>i}^{n} c_{\sqrt{\lambda_i^2+\lambda_j^2}} \int h_{ij}(\|r'_{ij}\|) e^{-\frac{\|r_{ij}-r'_{ij}\|^2}{\lambda_i^2+\lambda_j^2}} dr'_{ij}$$

$$= \sum_{i=1,j>i}^{n} < h_{ij} >_{\sqrt{\lambda_i^2+\lambda_j^2}} (\|r_{ij}\|), \tag{95}$$

which completes the proof. □

The anisotropic transform determines for any initial solution a unique solution function $x(\lambda)$ for the transformed functions, and therefore can also be used as a continuation process for optimization, more general and powerful than the isotropic transform. We state these results in Theorem 7 and 8. The details for the proof are quite similar to those for Theorem 2 and 3. So we will not present them.

Parallel to Assumptions 1, 2 and 3 for Theorem 2 and 3, we make for Theorem 7 and 8 the following assumptions:

Assumption 4 *The objective function f is twice continuously differentiable, and the transformation $<>_\Lambda$ is well defined for the function as well as its derivatives.*

Assumption 5 *Let g be the gradient of f, and Ψ an operator,*

$$\Psi = (\frac{\partial^2}{\partial x_1^2}, \frac{\partial^2}{\partial x_2^2}, \ldots). \tag{96}$$

Then the operation Ψ can be applied to g, and Ψg is a matrix with

$$(\Psi g)_{ij} = \Psi_j g_i = \frac{\partial^2 g_i}{\partial x_j^2}. \tag{97}$$

The transformation $<>_\Lambda$ is well defined for all derivatives involved in Ψg. Also, $\Psi g(x)$ is uniformly bounded and satisfies a Lipschitz condition:

$$\|\Psi g(x_1) - \Psi g(x_2)\| \leq L\ \|x_1 - x_2\| \qquad \forall x_1,\ x_2. \tag{98}$$

Assumption 6 *The transformation $< \nabla^2 f >_\Lambda (x)$ is uniformly bounded away from 0 and also satisfies a Lipschitz condition:*

$$\| <\nabla^2 f>_\Lambda (x_1) - <\nabla^2 f>_\Lambda (x_2)\| \leq L\ \|x_1 - x_2\| \qquad \forall x_1,\ x_2. \tag{99}$$

Let S be a vector space, and for a positive vector $\lambda_0 < \infty$, let $\Gamma = \prod_i [0, \lambda_{0i}]$. Then we define function $h(\lambda, x)$ such that $\forall(\lambda, x) \in \Gamma \times S$,

$$h(\lambda, x) = <f>_\Lambda (x), \tag{100}$$

where λ is the diagonal vector of Λ, i.e., λ corresponds to Λ uniquely,

$$\Lambda = \mathbf{diag}(\lambda). \tag{101}$$

Theorem 7 *Let f be a given function and h be defined as in (100). Then under Assumptions 4 and 5, $h''_{x\lambda}(\lambda, x)$ exists and is uniformly bounded for all $\lambda \in \Gamma$ and $x \in S$, and also satisfies a Lipschitz condition in x:*

$$\|h''_{x\lambda}(\lambda, x_1) - h''_{x\lambda}(\lambda, x_2)\| \leq L \ \|x_1 - x_2\| \qquad \forall x_1, x_2 \in S. \tag{102}$$

In addition,

$$h''_{x\lambda}(\lambda, x)^T = \frac{\Lambda}{2} < \Psi g >_\Lambda^T (x). \tag{103}$$

Theorem 8 *Let f be a function for which Assumptions 4, 5 and 6 all hold. Then for any stationary point x_0 of $< f >_{\Lambda_0}$, $\Lambda_0 = \mathbf{diag}(\lambda_0)$, there is a continuous and differentiable function $x(\lambda)$, $\lambda \in \Gamma$, such that $x_0 = x(\lambda_0)$ and $\forall \lambda \in \Gamma$, $x(\lambda)$ is a stationary point of $<f>_\Lambda$. The function $x(\lambda)$ is also the unique solution of the initial value problem*

$$x'^{\,T} = -\frac{\Lambda}{2}[<\nabla^2 f>_\Lambda^{-1} (x) <\Psi g>_\Lambda (x)]^T \tag{104}$$

$$x_0 = x(\lambda_0). \tag{105}$$

7 Concluding Remarks

In this paper, we discussed a generalization of the effective energy transformation scheme for the global energy minimization applied to molecular conformation [3, 4, 14, 15]. Instead of applying the transformation only to the probability distributions of given functions, here we transform the functions directly, generalizing the scheme in [3, 4, 14, 15] to a broader class of functions. Theories for the transformation as a general continuation approach are developed. We showed that the proposed method can transform a given nonlinear objective function into a class of gradually deformed but "smoother" functions. A general continuation procedure can then be applied to these "smoother" or "easier" functions, to trace the solution back to the original function. Two types of transformation are defined: the isotropic and the anisotropic. We showed that both, although not applicable numerically to arbitrary functions because of the required high dimensional integration, can be applied to a large class of nonlinear partially separable functions, and especially to the energy functions for molecular conformation. The methods to compute exactly the transformation for these functions are given.

We believe that the proposed method provides a powerful and effective tool for global or robust local optimization. We can see this partially from the work in [3, 4], which can be viewed as a special application of the method. In [3, 4], the transformation method, combined with simulated annealing, was applied to the global energy minimization problem for molecular conformation. Very promising performance and results were observed, even if only simple algorithms and approximated transformation were implemented (see [3, 4]).

More numerical work will be done in our future research. We will implement more efficient and effective algorithms based on the theories presented in this paper. While the transformation can now be computed with formulas provided, tracing the solution curve can be done using more developed numerical methods. There are at least three choices for the implementation of the tracing procedures:

1. Using a general random search procedure to trace the changes of the global solution when the transformed function is gradually changed back to the original function.

2. Applying only local optimization procedures to each transformed function to trace a set of solution curves, and choose the best among all solutions obtained at the end.

3. Solving the initial value problems for a set of solution curves, and choose the best solution.

The first method is similar to what has be done in [3, 4] where a simulated annealing process was applied to the transformed functions. This method guarantees the convergence to the global solution with certain probability, but a large number of random trials usually are required to obtain the convergence. The second one is the most simple and efficient way, but the solution curves to be traced must be selected very cleverly, for otherwise the global solution will not be guaranteed. The third method provides more accurate and reliable ways to trace the solution curves. As we have shown in this paper, the curves are solutions to well defined initial value problems. So standard numerical methods can be used, such as the predictor-corrector method. The implementation of all three tracing procedures and the numerical comparison among them will be of great interest for the development of the most effective and efficient algorithms.

We are especially interested in applying these methods to the global energy minimization problems for molecular conformation. A set of test problems will be considered including the Lennard-Jones microcluster conformation problem, the distance geometry problem, and the conformation problems for relatively small sizes of proteins.

While searching for native structures of protein molecules is certainly very important, the proposed methods can also provide the information about the paths that solutions change along. Such information may contain great insights about how protein molecules change from arbitrary configurations to their native structures. Work on this subject will be very interesting to do for the protein studies as well.

Acknowledgements

This research was supported partially by the Cornell Theory Center, which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its

Corporate Research Institute. The author would like to thank Lizhi Liao, Michael Todd, Lloyd Trefethen and Wei Yuan for constructive suggestions. He would especially like to thank Tom Coleman for many discussions relating to this work and for his helpful comments and suggestions on the manuscript, and David Shalloway for many discussions on the protein folding problem as well as original effective energy transformation ideas.

References

[1] Charles L. Brooks III, Martin Karplus and B. Montgomery Pettitt [1988]. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics.* John Wiley & Sons, New York, NY.

[2] Richard H. Byrd, Elizabeth Eskow, Robert B. Schnabel and Sharon L. Smith [1991]. *Parallel Global Optimization: Numerical Methods, Dynamic Scheduling Methods, and Application to Molecular Configuration.* Technical Report CU-CS-553-91, Department of Computer Science, University of Colorado at Boulder, Boulder, CO.

[3] Thomas Coleman, David Shalloway and Zhijun Wu [1992]. *Isotropic Effective Energy Simulated Annealing Searches for Low Energy Molecular Cluster States.* Technical Report CTC92TR113, Advanced Computing Research Institute, Cornell University, Ithaca, NY.

[4] Thomas Coleman, David Shalloway and Zhijun Wu [1993]. *A Parallel Build-Up Algorithm for Global Energy Minimizations of Molecular Clusters Using Effective Energy Simulated Annealing.* Technical Report CTC93TR130, Advanced Computing Research Institute, Cornell University, Ithaca, NY.

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

[6] J. E. Dennis, Jr. and R. B. Schnabel [1983]. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ.

[7] Philip E. Gill, Walter Murray and Margaret H. Wright [1981]. *Practical Optimization.* Academic Press, London.

[8] Bruce A. Hendrickson [1991]. *The Molecular Problem: Determining Conformation from Pairwise Distances.* Ph.D. Thesis, Department of Computer Science, Cornell University, Ithaca, NY.

[9] J. Kostrowicki, L. Piela, Binny J. Cherayil and Harold A. Scheraga [1991]. *Performance of the Diffusion Equation Method in Searches for Optimum Structures of Clusters of Lennard-Jones Atoms.* Journal of Physical Chemistry 95, 4113-4119.

[10] Z. Li and Harold A. Scheraga [1987]. *Monte Carlo Approach to the Multiple-Minima Problem in Protein Folding.* Proceedings of National Academy of Science, USA, 84, 15-29.

[11] J. A. Northby [1987]. *Structure and Binding of Lennard-Jones Clusters: $13 \leq n \leq 147$.* Journal of Chemical Physics 87, 6166-6178.

[12] L. Piela, J. Kostrowicki, and Harold A. Scheraga [1989]. *The Multiple-Minima Problem in the Conformational Analysis of Molecules. Deformation of the Protein Energy Hypersurface by the Diffusion Equation Method.* Journal of Physical Chemistry 93, 3339-3346.

[13] Enrico O. Purisima and Harold A. Scheraga [1987]. *An Approach to the Multiple-minima Problem in Protein Folding by Relaxing Dimensionality.* Journal of Molecular Biology, 196, 697-709.

[14] David Shalloway [1992]. *Packet Annealing: A Deterministic Method for Global Minimization. Application to Molecular Conformation.* In C. Floudas and P. Pardalos, eds, Global Optimization, Princeton University Press, Princeton, NJ.

[15] David Shalloway [1992]. *Application of the Renormalization Group to Deterministic Global Minimization of Molecular Conformation Energy Functions.* Global Optimization 2, 281-311.

[16] L. T. Wille [1987]. *Minimum-Energy Configuration of Atomic Clusters: New Results Obtained by Simulated Annealing.* Chemical Physics Letters 133, 405-410.

[17] Guo-liang Xue [1992]. *Improvement on the Northby Algorithm for Molecular Conformation: Better Solutions.* Army High Performance Computing Research Center, University of Minnesota, MN.