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

Zhijun Wu[†]

Abstract. This paper discusses a generalization of the function transformation scheme used in [4, 5, 19, 20] for global energy minimization applied to the molecular conformation problem. A mathematical theory for the method as a special continuation approach to global optimization is established. We show that the method can transform a nonlinear objective function into a class of gradually deformed, but "smoother" or "easier" functions. An optimization procedure can then be applied to the new functions successively, to trace their solutions back to the original function. Two types of transformation are defined: isotropic and anisotropic. We show that both transformations can be applied to a large class of nonlinear partially separable functions including energy functions for molecular conformation. Methods to compute the transformation for these functions are given.

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. Current address: Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439

1 Introduction

We are interested in solving the global minimization problem for molecular conformation, especially protein folding.

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

Optimization approaches to the protein folding problem are 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. The structure with the lowest energy is presumed to be the most stable protein structure.

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 globally minimize a nonlinear function $f(x)$ for all $x \in \mathbf{S}$, i.e.,

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

where $\mathbf{S}$ is the set of all possible molecular structures. The objective function $f(x)$ is the energy function for the protein. An important energy function type considered in this paper is of the pairwise interaction 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. Most potential energy terms for proteins can be written in form (2) such as the potentials for nonbonded interactions, the bonded potentials of bond lengths, and the potentials for hydrogen bonding. A widely used expression for nonbonded interactions, for example, is of the following Lennard-Jones plus Coulomb form:

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

where ϵ_{ij}, σ_{ij}, ϵ, q_i and q_j are all physical constants (see [2]).

Problem (1) is very difficult to solve in general. The reasons are as follows: First, in theory even simple versions of the problem have been proved to be $\mathcal{NP}$-complete [9]. Second, in practice the objective function often contains 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, the required computation is unaffordable using general global optimization methods.

However, because of its great practical importance, Problem (1) has been studied intensively in many areas of computational science and optimization. New algorithms on both sequential and parallel machines have been developed; a variety of small to medium sizes of problems have been studied [3, 4, 5, 6, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22]. In recent efforts smoothing techniques are specifically designed for molecular conformation via global minimization. Examples include the diffusion equation method [11, 12, 15], the packet annealing method [19, 20], as well as the effective energy simulated annealing method [4, 5]. The basic idea behind these methods is to use special techniques to smooth a given energy function so that search for a global minimizer becomes more tractable. The methods usually use function transformation schemes to transform a given energy function into a class of new functions. A solution tracing procedure is then applied to the new functions to locate a solution for the original function.

In this paper, we discuss an important generalization of the effective energy transformation scheme introduced in [4, 5, 19, 20]. Instead of applying the transformation to the probability distribution function, we now transform the functions directly, generalizing the method to a broader class of functions. More importantly, with this generalization, a mathematical theory for the transformation as a special continuation approach to global optimization is established. We show that the method can transform a nonlinear objective function into a class of gradually deformed, but "smoother" or "easier" functions. An optimization procedure can then be applied to the new functions successively, to trace their solutions back to the original function. Two types of transformation are defined: isotropic and anisotropic. We show that both transformation types can be applied to a large class of nonlinear partially separable functions which includes typical energy functions for molecular conformation. Methods to compute the transformation

for these functions are given.

The paper is organized as follows: Section 2 introduces the basic approach, and describes the function transformation method. Section 3 studies the mathematical properties of the transformation as a special continuation process. Section 4 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 5. The transformation is extended to the anisotropic type in Section 6. The formulas to compute the transformation for molecular conformation energy functions are derived. Finally, Section 7 contains concluding remarks.

2 The Approach

In this section, we describe our function transformation idea which, in turn, defines our basic approach to global optimization.

Suppose that we have a "poorly-behaved" nonlinear function with many local minimizers. Because of "nonsmoothness," this type of function can be very hard to minimize either locally or globally. To overcome this difficulty, we suggest using a special technique to transform the objective function into a class of gradually deformed, but "smoother" or "easier" functions. An optimization procedure can then be applied to these new functions successively, to trace their solutions back to the original function.

To deform the function, we define a 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

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

Note that in contrast to the approaches in [4, 5, 19, 20], the transformation here applies directly to the given function instead of its probability distribution. This simplifies the transformation, and also makes it much easier to compute and analyze.

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

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

In light of (7), transformation (4) yields a function value for $< f >_\lambda$ at any x equal to the expectation for f sampled by a Gaussian distribution function centered at x.

For example, consider the following nonlinear function:

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

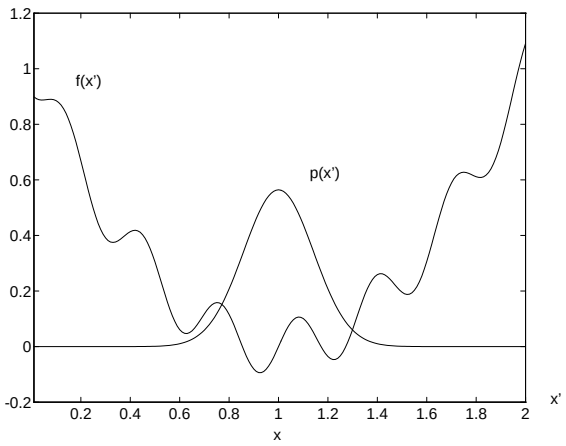
which is a quadratic function augmented with a "noise" function. The transformation for this function can be computed:

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

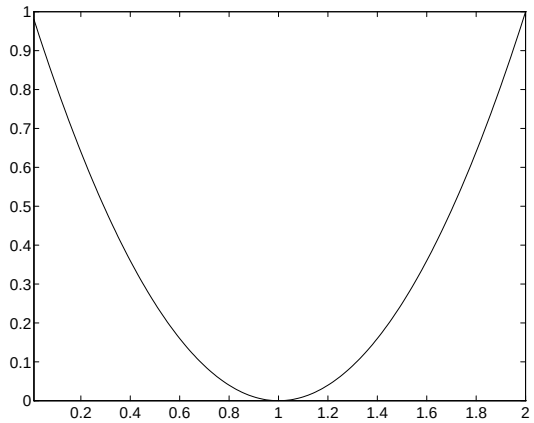
The function value $< f >_\lambda (x)$ for fixed x is equal to the integration with respect to the product of two functions, the original function $f(x')$ and the Gaussian distribution function $p(x') = \mathcal{C}_\lambda e^{-\|x-x'\|^2/\lambda^2}$ (Figure 1 (a)). The parameter λ determines the size 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 zero the transformed function is exactly the original function. For positive λ, the original function variations in small regions are averaged out, and the transformed function will become "smoother" (Figure 1 (b)).

Figure 2 shows how the function $< f >_\lambda$ in (9) behaves with increasing λ. Observe that when $\lambda = 0.0$ the function is 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".

Figure 3 illustrates what the transformation implies for optimization. A standard optimization procedure, a quasi-Newton method, is applied to the

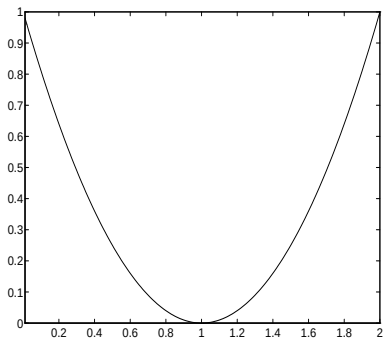


(a)

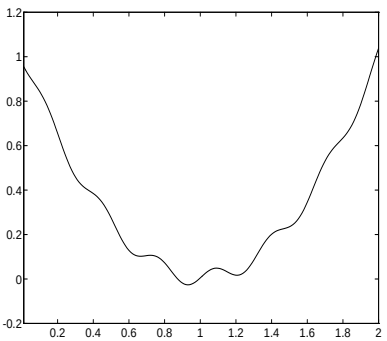


(b)

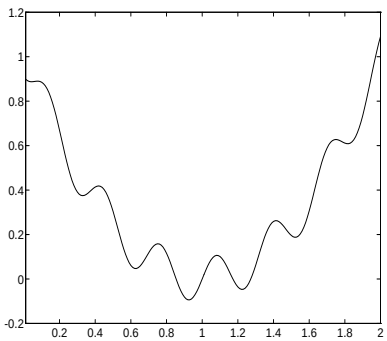
Figure 1: A one dimensional transformation example.



(a) $\lambda = 0.2$



(b) $\lambda = 0.1$



(c) $\lambda = 0.0$

Figure 2: A class of gradually deformed functions.

three functions in Figure 2. Figure 3 (a), (b), and (c) contain the corresponding solutions x^* obtained with different choices of initial guesses x°. Although globally convergent, the method may not find the right solution if the "noise" is large. So for the function in Figure 2 (c), the method converged to the right solution only when the initial guess was close enough to the solution. When the initial guess was far from the solution, the method failed to find the right solution (Figure 3 (c)). For the function in Figure 2 (b), although it is "smoother", the behavior of the method is essentially unchanged. However, for the function in Figure 2 (a), the method always converged to the right solution (Figure 3 (a)). If we apply the procedure to the functions in Figure 2 (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.

The experiment above suggests a general global optimization method: to optimize a difficult function, use the transformation technique to deform the function into a class of "smoother" or "easier" functions, and then apply an optimization procedure to the functions successively, to trace their solutions back to the original function.

3 Continuation

What is the difference between the suggested approach and 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. 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 variations gradually removed, but maintaining the overall function structure. Finally, if we apply an optimization procedure to a transformed function, the obtained solution usually is close to the solution for the original function. All these features are good for global optimization (also for robust local optimization), but are not necessarily the properties of conventional homotopies.

We show in the following that the proposed transformation is indeed a well

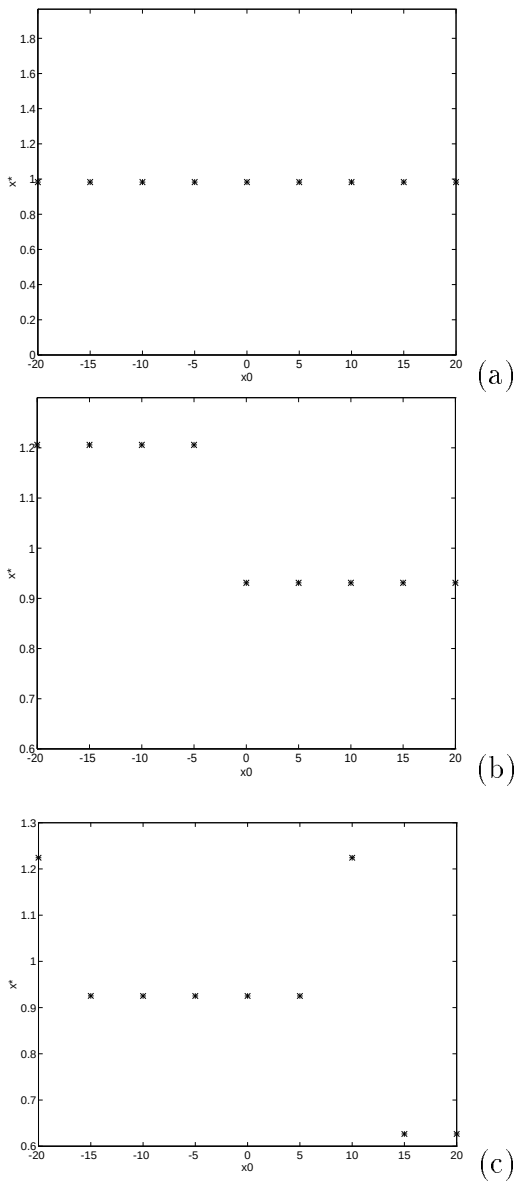


Figure 3: The solutions for the functions in Figure 2 obtained by the quasi-Newton's method with different initial guesses.

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 transformation (4) is well defined for the function as well as all 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 transformation (4) is well defined for all derivatives involved. 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 $<\nabla^2 f>_\lambda (x)$ satisfies a Lipschitz condition:*

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

and its inverse is uniformly bounded.

Note that to guarantee Assumptions 1 to 3, a sufficient condition on f is that f and its derivatives are all integrable in terms of parametrized integration (4).

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

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

$$\nabla <f>_\lambda (x) = <\nabla f>_\lambda (x) \tag{13}$$
$$\nabla^2 <f>_\lambda (x) = <\nabla^2 f>_\lambda (x). \tag{14}$$

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

$$\lim_{\lambda \to 0} < f >_\lambda (x) = f(x) \tag{15}$$

$$\lim_{\lambda \to 0} \nabla < f >_\lambda (x) = \nabla f(x) \tag{16}$$

$$\lim_{\lambda \to 0} \nabla^2 < f >_\lambda (x) = \nabla^2 f(x). \tag{17}$$

For 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 $\lambda \in \Gamma$ is

$$h'_x(\lambda, x) = 0. \tag{19}$$

Theorem 1 *Let $f : \mathbf{R} \to \mathbf{R}$ 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:*

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

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,

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

After differentiating (24) with respect to λ, it follows 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

$$\Delta_1 = \int x(x - x')g(x')p(\lambda, (x - x'))dx' \tag{26}$$

$$\Delta_2 = \int x'(x - x')g(x')p(\lambda, (x - x'))dx'. \tag{27}$$

It is easy to verify that

$$\Delta_1 = -\frac{\lambda^2}{2}x < g' >_\lambda (x) \tag{28}$$

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

Also note that

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

Therefore,

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

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

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

By Assumption 2, $<f'''>_\lambda (x)$ is well defined and uniformly bounded. So, $h''_{x\lambda}(\lambda, x)$ exists and is 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. □

Theorem 2 *Let $f : \mathbf{R}^n \to \mathbf{R}$ 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

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

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 of 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

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

Similar to the proof of 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. □

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*

$$x' = -\frac{\lambda}{2} < \nabla^2 f >_\lambda^{-1} (x) < \Delta g >_\lambda (x) \tag{48}$$

$$x_0 = x(\lambda_0). \tag{49}$$

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

$$h'_x(\lambda_0, x_0) = 0. \tag{50}$$

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

$$h_x^{'}(\lambda, x(\lambda)) = 0 \tag{51}$$

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 standard ordinary differential equation theory [10].

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. □

4 Smoothness

In Section 2 we illustrated 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

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

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

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

where ω is the frequency. So, we have

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

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

Also by (62), $\widehat{<f>_\lambda}(\omega)$ will be very small if ω is large and λ is fixed. This 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 increasingly smooth as λ increases. 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 $\forall \omega$ with $\|\omega\| > \delta$,*

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

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

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

□

From this theorem we learn that the relative size of $\widehat{< f >_\lambda}(\omega)$ can be made arbitrarily small for all ω with $\|\omega\|$ greater than a small value δ. Since δ is inversely proportional to λ, high frequency components are removed when λ is large.

5 Numerical Applicability

The definition of the transformation (4) involves high dimensional integration which cannot be computed in general (except perhaps by the Monte Carlo method which is not appropriate for our purposes because it is too expensive). So the transformation may not be applicable to arbitrary functions, at least numerically. However, this transformation is computationally feasible for a large class of nonlinear partially separable functions, and especially to typical molecular conformation and protein folding energy functions.

We state several useful properties of the transformation in the following:

First, for the sum of functions

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

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

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

Second, for the product of functions

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

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

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

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

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

where the 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 1-dimensional integration, the transformation for a generalized multilinear function can be computed using a standard quadrature rule.

In particular, 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 of this energy function is equal to the sum of the transformations of the pairwise energy functions. However the computation for the pairwise transformation still cannot be carried out directly, because there is still more than one variable in each term. 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 of 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

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

Proof: We show the case when $x_i \in \mathbf{R}$, $\forall i$. The general case can be proved similarly.

By the definition of $< f >_\lambda$, in form (5), for any x,

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

where c_λ is such that

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

Note that for $x_i \in R$, $c_\lambda = 1/(\lambda\sqrt{\pi})$. Make the following variable transformation:

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

Then for $x_i \in R$, $dx_i dx_j = \frac{1}{2} dr_{ij} d\bar{r}_{ij}$. Now it is easy to verify that

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

□

The integral for $< h_{ij} >_{\sqrt{2}\lambda} (\|r_{ij}\|)$ involves only variable r_{ij}, and can be computed with a standard numerical integration technique; therefore, the transformation $< f >_\lambda (x)$ can be computed in this fashion.

Note that the integral for $< f >_\lambda (x)$ must exist, for otherwise the transformation will have no definition. In practice, if the integral for a given f does not exist, a modified function may need to be considered instead. For example, the energy function given in (3) cannot be integrated directly because the function goes to infinity as r_{ij} goes to zero. Usually, this can be cured by replacing the function for small r_{ij} with finite interpolation (see [4, 12, 19]).

Note also that the result in Theorem 5 applies only to energy functions that can be formulated in form (71). Most potential energy terms for proteins can be expressed as pairwise forms, e.g, the Lennard-Jones potential, the electrostatic potential, the interaction potential for bonded atoms, etc. [2, 17]. However, protein energy functions do contain terms that are not pairwise distance functions, e.g., the torsional potential usually is given as a function of the dihedral angle. To transform this type of function, special approximation techniques may need to be used. We will not address this issue in this work.

6 Anisotropic vs. Isotropic

The transformation we have discussed so far is of the 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 a more accurate overall structure of the function. For this purpose, we can define a more general transformation, called the anisotropic transformation.

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

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

Note also that in this definition,

$$\|x/\Lambda\|^2 = \sum_{i=1}^{n} \|x_i\|^2/\lambda_i^2. \tag{82}$$

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

Many of the important properties of the isotropic transformation carry over to the anisotropic case. We state these properties in the following:

First, for the sum of the functions

$$f = \sum_i f_i, \tag{83}$$

we have

$$< f >_\Lambda = \sum_i < f_i >_\Lambda. \tag{84}$$

Second, for the product of the functions

$$g = \prod_i g_i, \tag{85}$$

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

$$< g >_\Lambda = \prod_i < g_i >_{\Lambda_i}, \tag{86}$$

where Λ_i's are small diagonal matrices. If g_i is a function of j variables $x_1^i, \ldots, x_j^i$, $\Lambda_i = \mathbf{diag}([\lambda_1^i, \ldots, \lambda_j^i])$ for j positive numbers $\lambda_1^i, \ldots, \lambda_j^i$.

Third, for the generalized multilinear functions,

$$f = \sum_i \prod_j g_j^i, \tag{87}$$

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

$$< f >_\Lambda = \sum_i \prod_j < g_j^i >_{\lambda_j}. \tag{88}$$

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 of f can be computed using the formula*

$$< f >_\Lambda (x) = \sum_{i=1, j>i}^{n} < h_{ij} >_{\sqrt{\lambda_i^2+\lambda_j^2}} (\|r_{ij}\|) \tag{89}$$

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

$$< h_{ij} >_{\sqrt{\lambda_i^2+\lambda_j^2}} (\|r_{ij}\|) = c_{\sqrt{\lambda_i^2+\lambda_j^2}} \int h_{ij}(\|r_{ij} - r'_{ij}\|) e^{-\frac{\|r'_{ij}\|^2}{\lambda_i^2+\lambda_j^2}} dr'_{ij}. \tag{90}$$

Proof: By the definition of $< f >_\Lambda$, in form (79), for any x,

$$\begin{aligned} & < f >_\Lambda (x) \\ = & \sum_{i=1, j>i}^{n} < h_{ij} >_\Lambda (\|x_i - x_j\|) \\ = & \sum_{i=1, j>i}^{n} c_{\lambda_i} c_{\lambda_j} \int h_{ij}(\|(x_i - x'_i) - (x_j - x'_j)\|) e^{-(\frac{\|x'_i\|^2}{\lambda_i^2} + \frac{\|x'_j\|^2}{\lambda_j^2})} dx'_i dx'_j, \end{aligned} \tag{91}$$

where c_{λ_i} is such that

$$c_{\lambda_i} \int e^{-\|x_i\|^2/\lambda_i^2} dx_i = 1, \qquad \forall i. \tag{92}$$

Make the following variable transformation:

$$r_{ij} = x_i - x_j, \qquad \bar{r}_{ij} = \lambda_j^2 x_i + \lambda_i^2 x_j. \tag{93}$$

Then we have

$$\frac{x_i}{\lambda_i} = \frac{1}{\lambda_i^2+\lambda_j^2}(\frac{\bar{r}_{ij}}{\lambda_i}+\lambda_i r_{ij}) \quad (94)$$

$$\frac{x_j}{\lambda_j} = \frac{1}{\lambda_i^2+\lambda_j^2}(\frac{\bar{r}_{ij}}{\lambda_j}-\lambda_j r_{ij}). \quad (95)$$

Using these relations we can verify that

$$\begin{aligned}
& <f>_\Lambda (x) \\
= & \sum_{i=1,j>i}^{n} c_{\lambda_i} c_{\lambda_j} \int h_{ij}(\|(x_i - x_i') - (x_j - x_j')\|) e^{-(\frac{\|x_i'\|^2}{\lambda_i^2}+\frac{\|x_j'\|^2}{\lambda_j^2})} dx_i' dx_j' \\
= & \sum_{i=1,j>i}^{n} c_{\sqrt{\lambda_i^2+\lambda_j^2}} \int h_{ij}(\|r_{ij} - r_{ij}'\|) e^{-\frac{\|r_{ij}'\|^2}{\lambda_i^2+\lambda_j^2}} dr_{ij}' \\
= & \sum_{i=1,j>i}^{n} <h_{ij}>_{\sqrt{\lambda_i^2+\lambda_j^2}} (\|r_{ij}\|), \quad (96)
\end{aligned}$$

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 the following assumptions:

Assumption 4 *The objective function f is twice continuously differentiable, and transformation (78) is well defined for the function as well as its derivatives.*

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

$$\Psi = (\frac{\partial^2}{\partial x_1^2}, \frac{\partial^2}{\partial x_2^2}, \ldots). \quad (97)$$

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

$$(\Psi g)_{ij} = \Psi_j g_i = \frac{\partial^2 g_i}{\partial x_j^2}. \tag{98}$$

Transformation (78) is well defined for all derivatives involved in Ψg. Also, $\Psi g(x)$ is uniformly bounded and satisfies a Lipschitz condition:

$$\|\Psi g(x_1) - \Psi g(x_2)\| \leq L \ \|x_1 - x_2\| \qquad \forall x_1, \ x_2. \tag{99}$$

Assumption 6 *The transformation $<\nabla^2 f>_\Lambda (x)$ satisfies a Lipschitz condition:*

$$\| <\nabla^2 f>_\Lambda (x_1) - <\nabla^2 f>_\Lambda (x_2)\| \leq L \ \|x_1 - x_2\| \qquad \forall x_1, \ x_2, \tag{100}$$

and its inverse is uniformly bounded.

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 $\forall (\lambda, x) \in \Gamma \times S$,

$$h(\lambda, x) = <f>_\Lambda (x), \tag{101}$$

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

$$\Lambda = \mathbf{diag}(\lambda). \tag{102}$$

Theorem 7 *Let f be a given function and h be defined as in (101). 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:*

$$\|h''_{x\lambda}(\lambda, x_1) - h''_{x\lambda}(\lambda, x_2)\| \leq L \ \|x_1 - x_2\| \qquad \forall x_1, x_2 \in S. \tag{103}$$

In addition,

$$h''_{x\lambda}(\lambda, x)^T = \frac{\Lambda}{2} < \Psi g >^T_\Lambda (x). \tag{104}$$

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*

$$x'^{\,T} = -\frac{\Lambda}{2}[< \nabla^2 f >_{\Lambda}^{-1} (x) < \Psi g >_{\Lambda} (x)]^T \tag{105}$$

$$x_0 = x(\lambda_0). \tag{106}$$

7 Concluding Remarks

In this paper, we have discussed a generalization of the effective energy transformation scheme used in [4, 5, 19, 20] for the global energy minimization applied to molecular conformation. Instead of applying the transformation to the probability distribution, here we transform the functions directly, generalizing the scheme in [4, 5, 19, 20] to a broader class of functions. A mathematical theory for the transformation as a special continuation approach to global optimization is established. We have established that the proposed method transforms a given nonlinear objective function into a class of gradually deformed, but "smoother" or "easier" functions. A continuation procedure can then be applied to these "smoother" or "easier" functions, to trace their solutions back to the original function. Two types of transformation are defined: isotropic and anisotropic. We have demonstrated that both transformation types can be applied to a large sub-class of nonlinear partially separable functions, and in particular, the energy functions for molecular conformation. Methods to compute 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 [4, 5], which can be viewed as a special application of the method. In [4, 5], the transformation method, combined with simulated annealing, was applied to the global energy minimization problem for molecular conformation. Promising results were observed even if only simple algorithms and approximated transformation were implemented.

More numerical work will be done in our future research. We will implement a group of algorithms based on the theory presented in this paper. While the transformation can now be computed with provided formulas, tracing the solution curve can be carried out using advanced numerical methods. There are at least three choices for the implementation of the tracing procedures:

1. Use 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. Apply only local optimization procedures to each transformed function to trace a set of solution curves, and choose the best among all solutions obtained.

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

The first method is similar to the approach in [4, 5] where a simulated annealing procedure was applied to the transformed functions. This method converges to the global solution with certain probability, but a large number of random trials usually are required to obtain the convergence. The second approach is the most simple and efficient method, but the solution curves to be traced must be selected cleverly, for otherwise the global solution will not be guaranteed. The third method provides a more accurate and reliable way 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 IVP-methods can be used, e.g., predictor-corrector methods [1]. The implementation of all three tracing procedures and the numerical comparison among them will be of great interest for the further development of the algorithms.

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

While searching for native structures of protein molecules is certainly very important, the proposed methods can also provide information about

the paths that solutions follow. Such information may contain insights about how protein molecules change from arbitrary configurations to their native structures.

Acknowledgements

This research was supported partially by the Cornell Theory Center, which receives funding from members of its Corporate Research Institute, the National Science Foundation (NSF), the Advanced Research Projects Agency (ARPA), the National Institutes of Health (NIH), New York State, and IBM Corporation.

The author would like to thank Lizhi Liao, Michael Todd, Lloyd Trefethen, and Wei Yuan for constructive suggestions. He would especially like to thank Thomas 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 the original effective energy transformation ideas.

Finally, the author appreciates the associate editor Robert Schnabel and the anonymous referee for their carefully reading the paper and for their very helpful suggestions.

References

[1] Eugene L. Allgower and Kurt Georg [1990]. *Numerical Continuation Methods.* Springer-Verlag, New York, NY.

[2] 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.

[3] 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.

[4] Thomas F. Coleman, David Shalloway and Zhijun Wu [1993]. *Isotropic Effective Energy Simulated Annealing Searches for Low Energy Molecular Cluster States.* Computational Optimization and Applications, 2, 145-170, 1993.

[5] Thomas F. Coleman, David Shalloway and Zhijun Wu [1994]. *A Parallel Build-Up Algorithm for Global Energy Minimizations of Molecular Clusters Using Effective Energy Simulated Annealing.* Journal of Global Optimization, 4, 171-185, 1994.

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

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

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

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

[10] David Kincaid and Ward Cheney [1991]. *Numerical Analysis.* Brooks/Cole Publishing Company, Pacific Grove, CA.

[11] J. Kostrowicki and L. Piela [1991]. *Diffusion Equation Method of Global Minimization: Performance for Standard Test Functions.* Journal of Optimization Theory and Applications 69, No. 2, 269-284.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] Harold A. Scheraga [1993]. *ECEPP/3 Empirical Conformation Energy Program for Peptides.* Department of Chemistry, Cornell University, Ithaca, NY.

[18] Tamar Schlick [1987]. *Modeling and Minimization Techniques for Predicting Three-Dimensional Structures of Large Biological Molecules.* Ph.D. Thesis, Courant Insititute of Mathematical Sciences, New York University, New York, NY.

[19] 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.

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

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

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