Journal of Global Optimization (2006) 34: 369–398
DOI 10.1007/s10898-005-4385-4

A Smoothing Method of Global Optimization that Preserves Global Minima

MARK S.K. LAU[1,*] and C.P. KWONG[2]
[1]*Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: sklau@acae.cuhk.edu.hk)*
[2]*Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: cpkwong@acae.cuhk.edu.hk)*

(Received 14 January 2004; accepted in revised form 16 March 2005)

Abstract. A new smoothing method of global optimization is proposed in the present paper, which prevents shifting of global minima. In this method, smoothed functions are solutions of a heat diffusion equation with external heat source. The source helps to control the diffusion such that a global minimum of the smoothed function is again a global minimum of the cost function. This property, and the existence and uniqueness of the solution are proved using results in theory of viscosity solutions. Moreover, we devise an iterative equation by which smoothed functions can be obtained analytically for a class of cost functions. The effectiveness and potential of our method are then demonstrated with some experimental results.

Key words: diffusion equation, global optimization, smoothing method, viscosity solution

1. Introduction

Various approaches have been proposed to tackle global optimization problems. Among them, smoothing methods are based on the intuition of smoothing out local minima of cost functions. The methods appeal to us, as they provide means of alleviating the difficulty of applying local search in global optimization.

The class of smoothing methods includes the diffusion equation method [19,21,26], the effective energy transformation method [31], and the others appearing in the contributions [7,24,25,27–29]. These methods have been applied mainly to optimum structures prediction of proteins and Lennard–Jones clusters; the first method has been found successful also in solving some standard test problems [18].

A smoothing method attempts to evolve the cost function into a (smoothed) function possessing only one minimum. This minimum can then be sought effectively by local search. Because the minimum may have shifted away from the global minimum of the cost function, a reversing procedure [26] has to be employed. However, for some cost functions, the shift can be so severe that the procedure ends up reaching merely a local

minimum. As there exist no useful conditions that would ensure convergence to the global minimum, smoothing methods are hindered by this difficulty.

A possible remedy is to ensure that the global minimum does not shift during evolution. The method proposed here uses a more sophisticated governing equation, under which the evolution preserves global minima. As we will show in Section 3, global minima of the smoothed and cost functions always coincide. As a result, the above-mentioned difficulty does not exist in our method.

Let f be the cost function of an unconstrained global optimization problem. The smoothed function in our method is defined as $u(t, \cdot)$, where u is a real-valued function on $[0, T) \times \mathbb{R}^n$ satisfying the governing equation

$$\frac{\partial u}{\partial t}(t, x) = \Delta u(t, x) - \max\{0, \Delta u(t, x)\}, \quad 0 < t < T, \tag{1}$$

with $u(0, \cdot) = f$. In this paper Δu denotes the Laplacian of u with respect to the "spatial variables". This equation models heat diffusion processes with the external heat source $-\max\{0, \Delta u\}$. This term controls diffusion processes by weakening the smoothing effect around minima of the smoothed functions. As a result, global minima of the cost function survive through the evolution. More precisely for $t_1 < t_2$ any global minimum of $u(t_1, \cdot)$ is again a global minimum of $u(t_2, \cdot)$. A more detailed discussion on the source term will appear in Section 2.

Another key issue in this paper regards computation of smoothed functions. To compute the functions, we need to solve the initial-value problem associated with the governing equation (1). The analytical solution is not pursued in this paper due to difficulties brought by the nonlinearity of the equation. Unfortunately, widely accepted solution schemes such as the finite difference method, collocation method and finite element method are infeasible. To approximate derivatives, these methods require discretization, i.e., evaluating the cost function on grid points. In case of high dimensional problems, they would render impotent as the number of grid points and function evaluations grow exponentially.

Under the circumstances, we devise a new scheme which does not require discretization. The central idea is to replace the Laplacian in the governing equation by an explicit expression. This leads to a simple iterative equation from which we can compute smoothed functions without function evaluation for cost functions of certain classes, such as linear combinations of Gaussian functions.

The potential and feasibility of the proposed method are then examined by experiments. The cost functions in the experiments are linear combinations of Gaussian functions. The results demonstrate that the method

is capable of smoothing out cost functions and also preserving global minima.

We remark that the proposed method is still in its infancy, and the objective here is to establish both theoretical and practical foundations for future developments. The paper is organized as follows : In Section 2, the role played by the source term is elaborated. Section 3 considers some theoretical aspects of the method, presenting our uniqueness and existence result and the global minima preserving property. The property is then illustrated by several simulation examples. In Section 4 a new scheme for computing smoothed functions is presented; a few issues associated with the scheme are also discussed. In Section 5 the potential of the proposed optimization method is demonstrated with experimental results.

2. Heat Diffusion with Source

Equation (1) is a special case of a more general equation:

$$\frac{\partial u}{\partial t}(t,x) = \Delta u(t,x) + S(t,x), \tag{2}$$

where S is a real function. This equation models many physical phenomena. The one described below, which appears in many standard textbooks, is probably the most well-known.

Suppose that there is an infinitely long metal rod with initial temperature distribution u_0 as shown in Figure 1. Assume that the rod is made of pure metal, perfectly insulated, and heated continuously over the time interval $(0, T)$ by a candle flame. It can be shown that if u is the solution of (2) with $u(0,\cdot) = u_0$, then $u(t,x)$ is equal to the temperature of the rod at (t,x). In this example $S(t,x)$ represents the temperature change at (t,x) induced

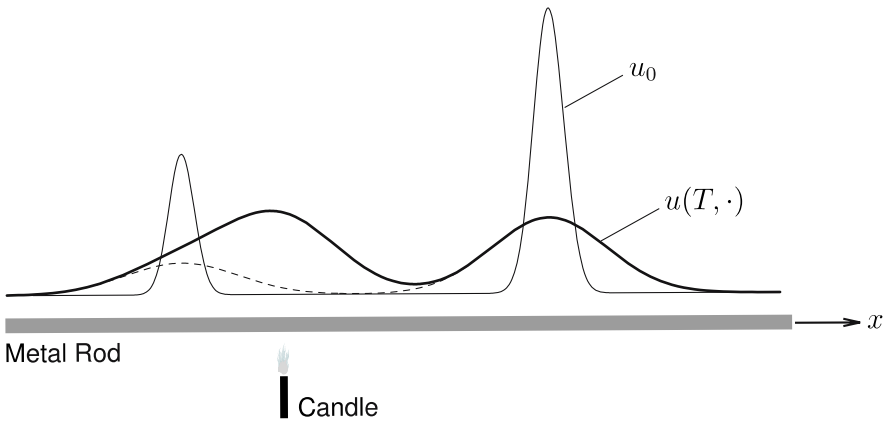


Figure 1. Controlling a heat diffusion process using a candle flame as external heat source.

by the external energy source, i.e., the candle flame. Clearly the final temperature distribution $u(T, \cdot)$ will be very different from the case when no external heat source has been applied (dashed line).

The role of the source term would be apparent if we think of the metal rod as a system[1] and $u(t, \cdot)$ as its state at time t. In this regard S becomes a control input to the system. It is suggested that by applying a suitable control, one could readily drive $u(t, \cdot)$ to a desired final state [12].

Our desired state is a function which has the same global minimum as that of the initial state (cost function). Therefore we define

$$S(t, x) = -\max\{0, \Delta u(t, x)\},$$

as given in (1). This source term counterbalances diffusion effects around minima of u. At a point close to a minimum, Δu is positive and hence the left hand side of (2) vanishes. As a result u does not evolve at that point. The minimum sustains through the smoothing for a longer period than the maxima do, and only global minima will survive eventually.

3. Theoretical Study

In this section, we study some properties of the initial-value problem associated with (1). First of all, we give a more precise problem formulation and define the terms which will appear frequently. This is followed by a quick overview of theory of viscosity solutions, which will provide us with the machinery for proving the properties. Finally we give several simulation examples to demonstrate the properties.

3.1. PROBLEM FORMULATIONS

The proposed optimization method tackles **unconstrained** global optimization problems. Let $f: \mathbb{R}^n \to \mathbb{R}$ be the cost function. The problem is to find a point $x^* \in \mathbb{R}^n$ such that $f(x^*) \leqslant f(x)$ for all $x \in \mathbb{R}^n$. We say that x^* is a **global minimum** of f if the inequality is satisfied; x^* is called a **minimum** of f if there exists an $\epsilon > 0$ such that $f(x^*) \leqslant f(x)$ for all $x \in \mathbb{R}^n$ with $\|x - x^*\| < \epsilon$, where $\|\cdot\|$ denotes the Euclidean distance. The existence of a minimum in an open set is not guaranteed in general [16], it is sufficient to assume that f has at least one minimum in $\mathbb{R}^n$.

To tackle the optimization problems, our method does not minimize the cost function directly. Instead, it first transforms f into a **smoothed function** $u(t, \cdot)$ for $0 < t < T$. Here u is the solution of the initial-value problem

[1] One may consider an one-variable function as a vector with infinite number of entries. Such a system is therefore referred to as an infinite dimensional system.

associated with the following equation and initial condition:

$$\begin{aligned} &\text{(GE)} \quad u_t(t,x) = \min\{0, \Delta u(t,x)\}, \quad (t,x) \in (0,T) \times \mathbb{R}^n \\ &\text{(IC)} \quad u(0,x) = f(x), \quad x \in \mathbb{R}^n, \end{aligned} \tag{3}$$

where u_t and Δu denote, respectively, the derivative $\partial u/\partial t$ and Laplacian

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}, \quad x = (x_1, x_2, \ldots, x_n)^{\mathsf{T}}.$$

We refer to this problem as a **smoothing problem**.

By the **initial function** we mean the function f or $u(0,\cdot)$. Initial functions are assumed to be continuous on $\mathbb{R}^n$ with at least one minimum. However, as we will see in Proposition 3.3, additional assumptions are needed for our main results to hold. The governing equation and initial condition are labeled as (GE) and (IC). The right hand sides of (GE) and (1) are indeed identical because $\Delta u - \max\{0, \Delta u\} = \min\{0, \Delta u\}$.

The following notations will be used subsequently: If X is a symmetric matrix, then $\mathrm{tr}(X)$ and $\|X\|$ denote its trace and operator norm, respectively. Moreover Du and D^2u denote, respectively, the gradient vector and Hessian matrix with respect to the spatial variables $x_1, x_2, \ldots, x_n$. The meanings of Δu and $\mathrm{tr}(D^2u)$ are clearly equivalent.

3.2. THEORY OF VISCOSITY SOLUTIONS

The aim of this subsection is to give a basic concept of viscosity solutions, and to prepare the reader for the theoretical studies. A formal definition of the solution and some useful results can be found in Appendix A.

Theory of viscosity solutions[2] considers a class of fully nonlinear partial differential equations (PDE) of second order. The notion was proposed to study problems that do not have differentiable solutions [9] (Examples can be found in [8,10,11]). In order to investigate those problems, one has to relax the requirements for a solution, and seek for a weaker notion. In this regard, definitions of viscosity solution (such as shown in Definition A.1) do not require differentiability, and hence allow merely continuous functions to be the solutions. As a consequence a viscosity solution, say v, does not necessarily satisfy the PDE everywhere in the domain. But if v is twice continuously differentiable at a point, then it must satisfy the equation at that point. Therefore v will be qualified as a classical solution if it is twice continuously differentiable on the domain. So, viscosity solution can be considered as a weak version of classical solution.

[2]The name refers to the vanishing viscosity method used in proving an existence result.

The main references for this paper are the contributions by Crandall [8,11], Chen [4], Giga [15] and Barles [1]. The first two papers are self-contained expositions of basic theories, concentrating on results for problems with bounded domains. The last two consider existence and uniqueness results of parabolic equations on unbounded domains, such as $\mathbb{R}^n$.

As we will see later, the theory provides us not only very general existence and uniqueness theorems, but also a lot of useful results under mild conditions. This machinery greatly simplifies the proofs of our results. Moreover, by virtue of the properties of viscosity solutions, our method is suitable for both smooth and nonsmooth initial functions. It can tackle optimization problems with nonsmooth cost functions.

Since viscosity solutions are the primary ones for the current discussion, we will occasionally drop the term "viscosity". It follows that subsolutions, supersolutions and solutions are various viscosity solutions unless specified otherwise.

3.3. MAIN RESULTS

Before presenting the main results, let us recall a definition and a well-known property of convex envelopes. The property is that every global minimum of a function is again a global minimum of its convex envelope, and both functions have the same optimum values. This will be used in proving the global minima preserving property.

DEFINITION 3.1. (Convex envelope). Let f be a lower semicontinuous function on Ω, where $\Omega \subset \mathbb{R}^n$ is nonempty and convex. The **convex envelope** of f over Ω is a function $\underline{f}$ such that

1. $\underline{f}$ is convex on Ω;
2. $\underline{f}(x) \leqslant f(x)$ for every $x \in \Omega$;
3. if h is a convex function on Ω such that $h(x) \leqslant f(x)$ for every $x \in \Omega$, then $h(x) \leqslant \underline{f}(x)$ for every $x \in \Omega$.

THEOREM 3.2. *Let $\Omega \subset \mathbb{R}^n$ be a convex set, and assume that $f: \Omega \to \mathbb{R}$ has a least one minimum in Ω. If $\underline{f}$ is the convex envelope of f over Ω, then*

$$f^* := \min_{x \in \Omega} f(x) = \min_{x \in \Omega} \underline{f}(x),$$
$$\{x \in \Omega : f(x) = f^*\} \subset \{x \in \Omega : \underline{f}(x) = f^*\}.$$

Now we turn to our existence and uniqueness result, which will be proved using Proposition 3.3. Before stating the proposition, let us introduce a

notation: By a **modulus** we mean a continuous function m on $[0, \infty)$ which is monotonically increasing on $(0, \infty)$ with $m(0) = 0$.

PROPOSITION 3.3 (Comparison). *Let f be the initial function of a smoothing problem, which satisfies the following conditions:*

- (*U*1) *There exists a number $K_1 > 0$ such that $|f(x)| \leqslant K_1(\|x\| + 1)$ for every $x \in \mathbb{R}^n$;*
- (*U*2) *There exists a modulus m such that $f(x) - f(y) \leqslant m(\|x - y\|)$ for every $x, y \in \mathbb{R}^n$;*
- (*U*3) *There exists a number $K_2 > 0$ such that $f(x) - f(y) \leqslant K_2(\|x - y\| + 1)$ for every $x, y \in \mathbb{R}^n$.*

Let μ and ω be, respectively, subsolution and supersolution of the problem. If there exists a number $K > 0$ such that

$$\mu(t, x) \leqslant K(\|x\| + 1) \quad \textit{and} \quad \omega(t, x) \geqslant -K(\|x\| + 1)$$

for every $(t, x) \in (0, T) \times \mathbb{R}^n$, then $\mu \leqslant \omega$ on $(0, T) \times \mathbb{R}^n$.

This proposition is a special version of Theorem A.3 shown in Appendix A; the latter was given in [15] and considers a wider class of parabolic equations. Our equation (GE) in fact belongs to this class since the assumptions (Fl)–(F4) hold for

$$F(X) := -\min\{0, \operatorname{tr}(X)\}, \quad X \in S^n.$$

The first three are due to the continuity and monotonicity of F. The last one is justified because each entry of $X \in S^n$ with $\|X\| \leqslant R$ has magnitude not exceeding $\sqrt{n}R$, and hence $c_R \leqslant n\sqrt{n}R$.

Proof of Proposition 3.3. Let μ^* and ω_* be, respectively, the upper semicontinuous envelope of μ and lower semicontinuous envelope of ω. (Definitions of semicontinuous envelopes will be given in Appendix A.) From (14) and (15) we see that $\mu^*(0, \cdot) \leqslant f \leqslant \omega_*(0, \cdot)$ and hence $\mu^*(0, x) - \omega_*(0, y) \leqslant f(x) - f(y)$ for every $x, y \in \mathbb{R}^n$. The conditions (V2) and (V3) in Appendix A are therefore satisfied. The condition (V1) is also satisfied because of the assumptions on μ and ω. We then apply Theorem A.3 to prove the proposition. □

In Proposition 3.3 the "growth conditions" (U1)–(U3) are imposed to restrict the behavior of the solution at infinity. Indeed, for most PDE, comparison results fail to hold on unbounded domains without growth conditions. (The heat equation is an example [13].) Exceptional cases are discussed in [1].

Using Proposition 3.3 in conjunction with Perron's method, we can show that the smoothing problem has unique solution. This is a rather standard technique, and we will essentially follow the proofs of Theorem 3.1 of [1] and Theorem 4.1 of [11]. We therefore defer the proof to Appendix A without distracting the current discussions.

PROPOSITION 3.4 (Existence and uniqueness). *Let f be the initial function of a smoothing problem, which satisfies $(U1)$–$(U3)$. Then the problem has a unique viscosity solution u, which is continuous on $[0, T) \times \mathbb{R}^n$ and satisfies $|u(t, x)| \leqslant K(\|x\| + 1)$ for a number $K > 0$ and every $(t, x) \in (0, T) \times \mathbb{R}^n$.*

Our next task is to prove the global minima preserving property. This will be accomplished by applying the following propositions. They state that $u(t, \cdot)$ is monotonically decreasing but bounded below by the convex envelope of the initial function.

PROPOSITION 3.5. *Let f and u be, respectively, the initial function and viscosity solution of a smoothing problem. If f satisfies $(U1)$–$(U3)$, then $\underline{f}(x) \leqslant u(t, x) \leqslant f(x)$ for every $(t, x) \in (0, T) \times \mathbb{R}^n$, where $\underline{f}$ is the convex envelope of f over $\mathbb{R}^n$.*

Proof. Let $g(t, x) = f(x)$ and $\underline{g}(t, x) = \underline{f}(x)$ for each $(t, x) \in [0, T) \times \mathbb{R}^n$. Proposition A.4 implies that g and $\underline{g}$ are, respectively, viscosity supersolution and subsolution of the smoothing problem. Because $\underline{g}(t, x) \leqslant f(x) \leqslant K_1(\|x\| + 1)$ and $g(t, x) \geqslant -K_1(\|x\| + 1)$, where K_1 is given in (U1), Proposition 3.3 suggests that $\underline{f}(x) = \underline{g}(t, x) \leqslant u(t, x) \leqslant g(t, x) = f(x)$. □

PROPOSITION 3.6 (Monotonicity). *Let f and u be, respectively, the initial function and viscosity solution of a smoothing problem. If f satisfies $(U1)$–$(U3)$, then $u(t_2, x) \leqslant u(t_1, x)$ for any $0 \leqslant t_1 < t_2 < T$ and $x \in \mathbb{R}^n$.*

Proof. Let $v(t, x) = u(t + \tau, x)$ for $\tau > 0$ and every $(t, x) \in [0, T - \tau) \times \mathbb{R}^n$. As suggested by Proposition A.5 v is a subsolution of the problem on $(0, T - \tau) \times \mathbb{R}^n$. Because $v(t, x) \leqslant K(\|x\| + 1)$ for some $K > 0$, we invoke the comparison to show that $u(t + \tau, x) \leqslant u(t, x)$. The proposition is proved by putting $t_1 = t$ and $t_2 = t + \tau$. □

In fact $\underline{f}$ is also the convex envelope of $u(t, \cdot)$ over $\mathbb{R}^n$ for each fixed t. This can be verified by considering a convex function $h : \mathbb{R}^n \to \mathbb{R}$ such that $h \leqslant u(t, \cdot)$ on $\mathbb{R}^n$. Proposition 3.5 suggests that $h \leqslant f$ on $\mathbb{R}^n$, and therefore $h \leqslant \underline{f}$ on $\mathbb{R}^n$ by the definition of convex envelopes. This proves the assertion.

Given these ingredients, we are ready to prove the global minima preserving property. By combining Propositions 3.5 and 3.6, we obtain the inequality:

$$\underline{f}(x) \leqslant u(t_2, x) \leqslant u(t_1, x) \leqslant f(x) \tag{4}$$

for every $x \in \mathbb{R}^n$ whenever $0 \leqslant t_1 < t_2 < T$. This equation implies that any global minimum of $u(t_1, \cdot)$ is again a global minimum of $u(t_2, \cdot)$. This can be proved by contradiction as follows.

Suppose that x_1 is a global minimum of $u(t_1, \cdot)$ but not a global minimum of $u(t_2, \cdot)$. If x_2 is a global minimum of $u(t_2, \cdot)$, then $u(t_2, x_2) < u(t_2, x_1) \leqslant u(t_1, x_1)$. Since $\underline{f}$ is the convex envelope of both $u(t_1, \cdot)$ and $u(t_2, \cdot)$ over $\mathbb{R}^n$, they must have the same optimal values, i.e., $u(t_2, x_2) = u(t_1, x_1)$ in view of Theorem 3.2. This contradicts the last inequality. We have just proved the following proposition:

PROPOSITION 3.7 (Global Minima Preserving). *Let u be the viscosity solution of a smoothing problem whose initial function satisfies $(U1)$–$(U3)$. For $0 \leqslant t_1 < t_2 < T$, if x^* is a global minimum $u(t_1, \cdot)$, then it is also a global minimum of $u(t_2, \cdot)$ and $u(t_1, x^*) = u(t_2, x^*)$.*

Because of the dissipative property of (GE), the roughness of $u(t, \cdot)$ decreases with time. At the moment, we lack useful rules to determine if the function would eventually become unimodal. Nevertheless, computer simulations for bivariate functions indicate that minima in relatively "narrow" basins usually disappear rather quickly. It appears that if the initial function resembles a combination of a unimodal function and a low-amplitude "noisy" signal, then $u(t, \cdot)$ would probably become unimodal for t large enough.

Unfortunately, as we observe, smoothed functions usually evolve very slowly at saddle points, say x_s, with $\Delta u(t, x_s) \geqslant 0$. At those points, evolution slows down because $u_t(t, x_s) = \min\{0, \Delta u(t, x_s)\} = 0$. It may take quite a long time to smooth out the barrier at x_s.

This difficulty can be remedied by substituting $\lambda_{\min}(D^2 u)$ into (GE) for Δu, where $\lambda_{\min}(D^2 u)$ denotes the smallest eigenvalue of $D^2 u$. A similar idea is employed by the evolution equation proposed in [30]. The author showed that the viscosity solution of the equation under appropriate boundary conditions converges to the convex envelope of the initial function. However, the analytical solution remains unknown at the moment. The equation has little contribution to the current research unless a suitable solution method can be devised.

Remark 3.1. Let $Q \subset \mathbb{R}^n$ be bounded. By virtue of Theorem 4.1 of [4], our comparison result holds for the initial-boundary-value problem of

$$u_t = \min\{0, \Delta u\} \text{ on } (0, T) \times Q \quad \text{and } u = f \text{ in } (\{0\} \times Q) \cup ([0, T] \times \partial Q)$$

without the conditions (Ul)–(U3). Namely, if μ and ω are, respectively, viscosity subsolution and supersolution of the PDE satisfying $\mu \leqslant \omega$ in $(\{0\} \times Q) \cup ([0, T] \times \partial Q)$, then $\mu \leqslant \omega$ on $(0, T) \times Q$. Using this result one can prove that (4) and Proposition 3.7 hold for any initial function and $x, x^* \in Q$. The existence and uniqueness of the solution u follow directly from Theorem 4.4 of [4].

3.4. NUMERICAL EXAMPLES

We now demonstrate the theoretical results with several simulation examples, as shown in Figures 2 and 3. To do that, we approximate the values of the smoothed functions at a set of grid points using finite difference schemes. The grid point with the lowest function value will be considered as the global minimum. Such a point should be a good estimate for the exact global minimum if the grid is fine enough.

Because finite difference schemes assume bounded domains with fixed boundary values, we need to consider a domain "larger" than that of our interest. For example, suppose that we are interested in observing the smoothed functions on $[a, b]$. Then we solve $\phi_t = \min\{0, \Delta\phi\}$ on $(0, T) \times (\bar{a}, \bar{b})$ under the condition $\phi = f$ in $(\{0\} \times [\bar{a}, \bar{b}]) \cup ([0, T] \times \{\bar{a}, \bar{b}\})$. Here we choose $\bar{a} < a$ so small and $\bar{b} > b$ so large that $\phi \approx u$ on $(0, T) \times [a, b]$, where u is the solution of the smoothing problem. The same technique is used for bivariate functions.

We will propose another scheme in Section 4, which generates smoothed functions analytically. The scheme, however, is not suitable for this demonstration because it is still in its infancy and only applicable to certain classes of functions.

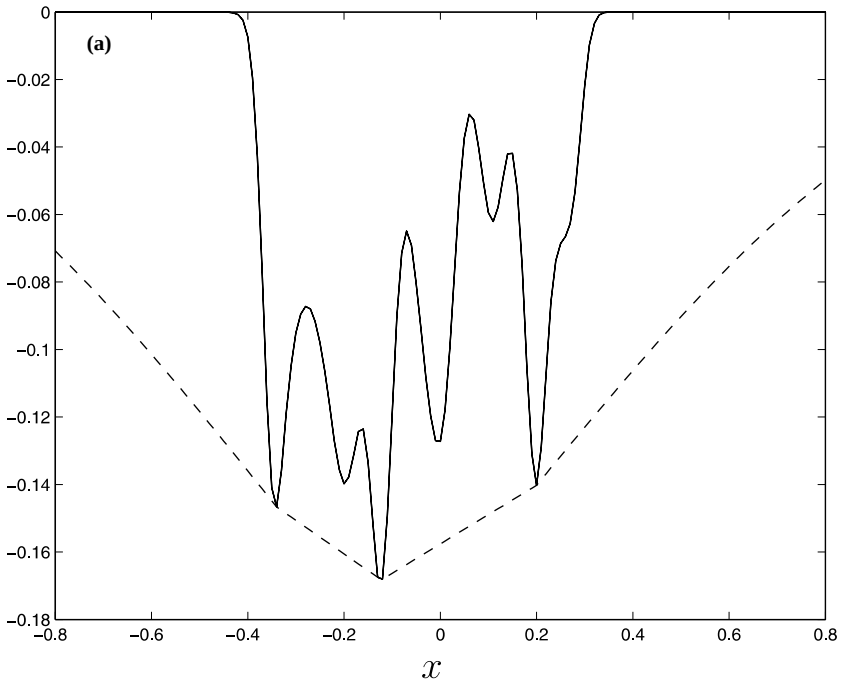
The initial function in the first example is a linear combination of Gaussian functions, which takes the form

$$f(x) = -\sum_{j=1}^{N} c_j \exp(-a_j \|x + y_j\|^2),$$

where $x, y_j \in \mathbb{R}$ and $a_j, c_j > 0$ for $j = 1, 2, \ldots, N$. This function and its smoothed function are displayed in Figure 2(a) using solid-lines and dashed-lines, respectively. The figure shows that f has six minima in $[-0.8, 0.8]$. All of them are smoothed out except the global minimum, which coincides with the global minimum of the smoothed function. Moreover, both functions have identical optimum values.

The initial function for the second example is the polynomial

$$f(x) = 162.01x^8 + 21.20x^7 - 221.57x^6 - 18.88x^5 + 96.24x^4 + 4.01x^3 - 13.44x^2 + 0.07x + 0.67.$$



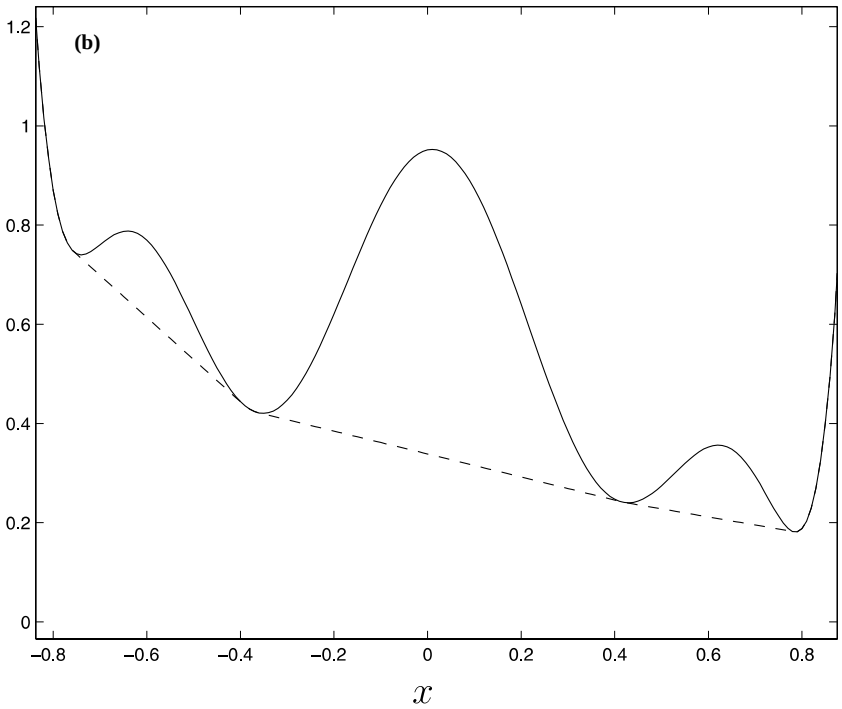


Figure 2. One-variable initial functions and their smoothed functions.

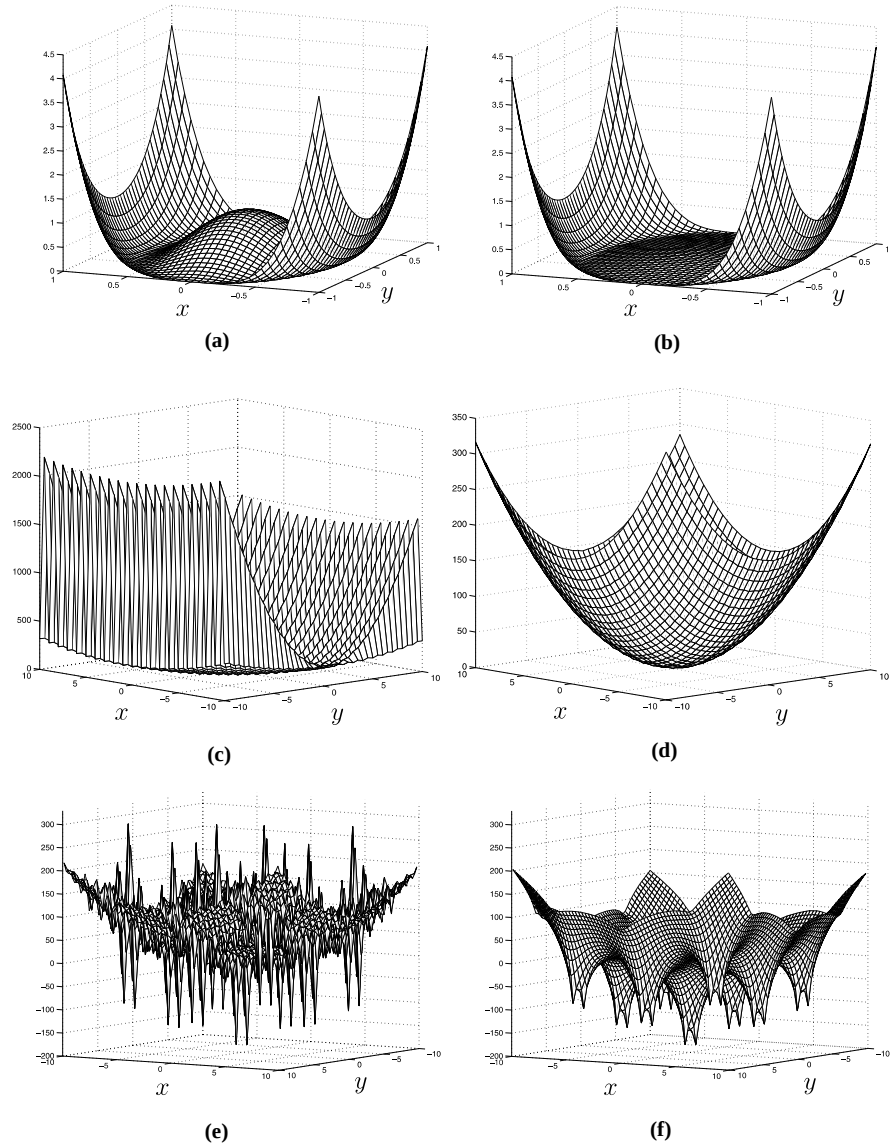


Figure 3. Two-variable initial functions and their smoothed functions.

This function and its smoothed function are plotted in Figure 2(b) using solid-lines and dashed-lines, respectively. Like the previous example, the global minimum of f is preserved while other minima disappear gradually. The following observation is worth mentioning. The smoothed function is very close to the convex envelope of the initial function. As we observe, the former keeps approaching the latter as the smoothing continues. It appears that the smoothed function converges uniformly to the convex envelope.

Next we consider three examples in which the initial functions depend on two variables. Namely, the polynomial $f(x, y) = (0.98 - x^2 - y^2)^2$, the sine-square function

$$f(x, y) = \frac{\pi}{2}\left\{10 \sin^2(\pi x) + (x-1)^2[1 + 10 \sin^2(\pi y)] + (y-1)^2\right\}$$

and the Shubert III function

$$f(x, y) = \left\{\sum_{j=1}^{5} j \cos[(j+1)x + 1]\right\}\left\{\sum_{j=1}^{5} j \cos[(j+1)y + 1]\right\} + (x + 1.42513)^2 + (y + 0.80032)^2.$$

The last two functions have been widely used for benchmarking global optimization algorithms (see [14, 23] for instances). The last function is moreover recognized as an important test function because of the large number of minima and the steep slope around the global minimum.

As seen in Figure 3(a) and (b), each global minimum of the polynomial coincide with a global minimum of the smoothed function. The set of all global minima of the polynomial, i.e., the circle $x^2 + y^2 = 0.98$, is actually contained in that of the smoothed function. This agrees with Proposition 3.7.

Shown in Figure 3(c), the sine-square function has around 100 minima in the domain $[-10, 10]^2$ as reported in [14]. They are all smoothed out except the global minimum, which is preserved as seen in Figure 3(d). The smoothed function is unimodal, having the same optimum value as that of the initial function. As we observe, the function is very close to the convex envelope of the initial function over $[-10, 10]^2$.

Let us turn to the Shubert III function shown in Figure 3(e). The result is less appealing compared with the previous examples. The smoothed function has more than one minimum as seen from 3(f). Because saddle points have formed, the evolution slows down. But the smoothing still yields a notable reduction on the number of minima: The initial function has 760 minima in the domain $[-10, 10]^2$ but only 18 of them remain. The figures show that the global minimum is preserved.

4. A Solution Method for Smoothing Problem

Because smoothed functions are solutions of smoothing problem, the effectiveness of the proposed optimization method largely relies on an efficient solution scheme. Unfortunately, the governing equation (GE) is a nonlinear PDE whose analytical solution is generally difficult to obtain.

Even worse, it seems to us that widely employed solution schemes, such as the finite difference method, the collocation method and the finite element method, are infeasible in the present application. This is because they require evaluating initial functions on grid points (or discretization) in order to approximate derivatives. In case of multi-variable problems, the schemes would render impotent as the number of grid points grows exponentially. Moreover, for many cost functions, a good estimate of the global minimum can be obtained merely by comparing the function values at the grid points. Therefore it would be redundant to carry out smoothing after discretization.

In this section a new solution method is proposed, which does not require discretization. The main idea is to replace the Laplacian $\Delta u(t, x)$ in (GE) by the difference equation

$$\Delta u(t, x) \approx \frac{1}{\delta t}[\mathcal{T}(\delta t)u(t, x) - u(t, x)], \quad (t, x) \in (0, T) \times \mathbb{R}^n, \tag{5}$$

where δt is the time step, and $\mathcal{T}(\delta t)$ is an operator to be introduced shortly. As we will see, a simple iterative scheme can be constructed based on this equation. Using this scheme, we can compute smoothed functions analytically (without discretization) whenever $\mathcal{T}(\delta t)u$ can be obtained explicitly.

4.1. DERIVATIONS

Let us first give a formal derivation of (5). Consider a function v which satisfies the heat equation with the initial condition:

$$\begin{aligned} &v_\tau(\tau, x) = \Delta v(\tau, x), \quad (\tau, x) \in (0, T) \times \mathbb{R}^n, \\ &v(0, x) = v_0(x), \quad x \in \mathbb{R}^n. \end{aligned} \tag{6}$$

Here we denote by τ the time variable, so that it will not be confused with that of smoothing problem. The solution is equal to [17]

$$v(\tau, x) = \mathcal{T}(\tau)v_0(x) \quad \text{for every } (\tau, x) \in [0, T) \times \mathbb{R}^n,$$

where $\mathcal{T}$ is called the **diffusion operator**, written as

$$\mathcal{T}(\tau)v_0(x) = \frac{1}{(2\sqrt{\pi\tau})^n} \int_{\mathbb{R}^n} v_0(y) e^{-\frac{1}{4\tau}\|x-y\|^2} dy.$$

The solution v is infinitely differentiable on $(0, T) \times \mathbb{R}^n$. So, we invoke Taylor's theorem to express $\Delta v(\tau, x)$ as

$$\begin{aligned} \Delta v(\tau, x) &= v_\tau(\tau, x) \\ &= \frac{v(\delta t + \tau, x) - v(\tau, x)}{\delta t} + O(\delta t) \quad \text{as } \delta t \to 0. \end{aligned}$$

Assume that v_0 is twice continuously differentiable on $\mathbb{R}^n$. Because v and Δv are continuous on $[0, T) \times \mathbb{R}^n$,

$$\begin{aligned}\Delta v_0(x) &= \lim_{\tau \to 0} \Delta v(\tau, x) \\ &= \lim_{\tau \to 0} \frac{v(\delta t + \tau, x) - v(\tau, x)}{\delta t} + O(\delta t) \\ &= \frac{v(\delta t, x) - v_0(x)}{\delta t} + O(\delta t) \quad \text{as} \quad \delta t \to 0.\end{aligned}$$

Finally, we set $v_0 = u(t, \cdot)$ on $\mathbb{R}^n$. Because $v(\delta t, x) = \mathcal{T}(\delta t) v_0(x) = \mathcal{T}(\delta t) u(t, x)$, we obtain the approximation (5), namely

$$\Delta u(t, x) = \frac{\mathcal{T}(\delta t) u(t, x) - u(t, x)}{\delta t} + O(\delta t) \quad \text{as} \quad \delta t \to 0.$$

Now substitute the above equation and the difference equation

$$u_t(t, x) = \frac{u(t + \delta t, x) - u(t, x)}{\delta t} + O(\delta t)$$

into (GE) to obtain

$$u(t + \delta t, x) = u(t, x) + \min\{0, \mathcal{T}(\delta t) u(t, x) - u(t, x)\} + O(\delta t^2) \tag{7}$$

as $\delta t \to 0$ for every $(t, x) \in (0, T) \times \mathbb{R}^n$. This equation prompts us to define the following iterative equation:

$$u_{k+1}(x) = u_k(x) + \min\{0, \mathcal{T}(\delta t) u_k(x) - u_k(x)\} \tag{8}$$

for $k = 0, 1, \ldots, T \delta t^{-1} - 2$. Equation (7) suggests that if $u_0 := f$ where f is the initial function, then u_k is an approximation of $u(k \delta t, \cdot)$.

Suppose that $u_0, u_1, \ldots, u_k$ are known. Then u_{k+1} can be computed using (8) without evaluating u_k if

$$\mathcal{T}(\delta t) u_k(x) = \frac{1}{\left(2\sqrt{\pi \delta t}\right)^n} \int_{\mathbb{R}^n} u_k(y) e^{-\frac{1}{4\delta t} \|x - y\|^2} dy \tag{9}$$

can be solved explicitly. However, this can hardly be achieved because u_k in the integrand is in the form $u_k = u_{k-1} + \min\{0, \mathcal{T}(\delta t) u_{k-1} - u_{k-1}\}$ and the zeros of $\mathcal{T}(\delta t) u_{k-1} - u_{k-1}$ are assumed to be unknown. This difficulty will be addressed in the next subsection.

4.2. APPLICATION TO SMOOTHING OF GAUSSIAN FUNCTIONS

As mentioned previously, the integral (9) is usually unable to be solved explicitly. Our remedy is to replace $\min\{0, \cdot\}$ in (8) with an approximating polynomial. As we will show later, if the initial function is a series of Gaussian functions, then u_k generated from the new equation is again a series of Gaussian functions. This ensures that the integral (9) can be solved explicitly for every k.

We remark that we have not imposed any strong assumptions on the initial function in deriving (8). The equation can thus be employed for other types of initial functions, provided that the integral (9) can be obtained or approximated efficiently.

4.2.1. *A Modified Iterative Equation*

By a **series of Gaussian functions** we mean a function defined by

$$f(x) = c_1 g_1(x) + c_2 g_2(x) + \cdots + c_N g_N(x)$$

for every $x \in \mathbb{R}^n$. We refer to N as the **length** of f, and to $-1 \leqslant c_i \leqslant 1$ as the **coefficients** of f. In this equation g_i are **Gaussian functions,** namely,

$$g_i(x) = \exp(a_i \|x + y_i\|^2),$$

where $y_i \in \mathbb{R}^n$ and a_i are negative numbers. Moreover, we denote by $G(n)$ the set of all series of Gaussian functions of n variables.

Suppose that the operator $\min\{0, \cdot\}$ in (8) is now replaced by a polynomial of rth order. The new equation is then in the following form:

$$\begin{aligned} u_{k+1} = q_0 + u_k + q_1[\mathcal{T}(\delta t)u_k - u_k] + q_2[\mathcal{T}(\delta t)u_k - u_k]^2 \\ + \cdots + q_r[\mathcal{T}(\delta t)u_k - u_k]^r. \end{aligned}$$

If $u_k \in G(n)$, then $\mathcal{T}(\delta t)u_k - u_k \in G(n)$. This is because $\mathcal{T}(\delta t)$ is a linear operator and $\mathcal{T}(\delta t)g_i(x) = b_i^{n/2} \exp(a_i b_i \|x + y_i\|^2)$ with $b_i = (1 - 4a_i \delta t)^{-1}$. Moreover, expanding $(g_1 + \cdots + g_N)^r$ results in another series of Gaussian functions. Thus we see that $u_{k+1} \in G(n)$ if $u_k \in G(n)$ for all k.

The polynomial we use is the Lagrange interpolating polynomial [3, 5, 6]

$$M_{p,\alpha}(\xi) := \sum_{j=0}^{2p+1} \left(\min\{0, \alpha\xi_j\} \prod_{\substack{k=0 \\ k \neq j}}^{2p+1} \frac{\xi - \alpha\xi_k}{\alpha\xi_j - \alpha\xi_k} \right).$$

This polynomial interpolates the data

$$(\alpha\xi_0, \min\{0, \alpha\xi_0\}), (\alpha\xi_1, \min\{0, \alpha\xi_1\}), \ldots, (\alpha\xi_{2p+1}, \min\{0, \alpha\xi_{2p+1}\}),$$

where $\alpha > 0$ is a parameter and ξ_j are the zeros of the Chebyshev polynomial of degree $2p+2$, namely

$$\xi_j = \cos\left[\frac{2j+1}{4(p+1)}\pi\right] \quad \text{for} \quad j = 0, 1, \ldots, 2p+1.$$

Instances of ξ_j and $M_{p,\alpha}$ are given in Table 1 and Figure 4 for illustration.

Note that $M_{p,\alpha}$ minimizes the maximum absolute error [2,3], i.e.,

$$\max_{\xi\in[-\alpha,\alpha]} |\min\{0, \xi\} - M_{p,\alpha}(\xi)| \leqslant \max_{\xi\in[-\alpha,\alpha]} |\min\{0, \xi\} - P(\xi)| \tag{10}$$

for every polynomial P of degree at most $2p+1$. In this sense, it is the best approximation over the interval $[-\alpha, \alpha]$. Nonetheless $|\min\{0, \xi\} - M_{p,\alpha}(\xi)|$ can be very large for $|\xi| > \alpha$. To achieve a high fidelity approximation, α should be chosen such that $|\mathcal{T}(\delta t)u_k(x) - u_k(x)| \leqslant \alpha$ for every $x \in \mathbb{R}^n$.

It is worth noting that $M_{p,\alpha}$ is of even degree. This is because we have used only the zeros of even-degree Chebyshev polynomials. In fact the zeros of the odd-degree counterparts include the origin, at which $\min\{0, \cdot\}$ is non-differentiable. This worsens the maximum error (10).

Finally, we replace $\min\{0, \cdot\}$ in (8) with M_{p,α_k} to obtain

$$u_{k+1}(x) = u_k(x) + M_{p,\alpha_k}[\mathcal{T}(\delta t)u_k(x) - u_k(x)]. \tag{11}$$

The parameter α_k is chosen such that $|\mathcal{T}(\delta t)u_k(x) - u_k(x)| \leqslant \alpha_k$ for every $k = 0, 1, \ldots, \mathcal{T}\delta^{-1} - 2$ and $x \in \mathbb{R}^n$. The selection of p and δt will be discussed at the end of this subsection.

Table 1. The polynomial $M_{p,\alpha}$ and zeros ξ_j of the Chebyshev polynomials of degree $2p+2$

p	$M_{p,\alpha}(\xi)$
1	$-0.1353/\alpha + 0.5\xi - 0.3827\alpha\xi^2$
2	$-0.0863/\alpha + 0.5\xi - 0.6609\alpha\xi^2 + 0.2526\alpha^3\xi^4$
3	$-0.0637/\alpha + 0.5\xi - 0.9204\alpha\xi^2 + 0.8479\alpha^3\xi^4 - 0.3662\alpha^5\xi^6$
4	$-0.0506/\alpha + 0.5\xi - 1.1728\alpha\xi^2 + 1.8948\alpha^3\xi^4 - 1.8732\alpha^5\xi^6 + 0.7031\alpha^7\xi^8$
p	ξ_j
1	$\pm 0.3826, \pm 0.9239$
2	$\pm 0.2588, \pm 0.7071, \pm 0.9659$
3	$\pm 0.1951, \pm 0.5556, \pm 0.8315, \pm 0.9808$
4	$\pm 0.1564, \pm 0.4540, \pm 0.7071, \pm 0.8910, \pm 0.9877$

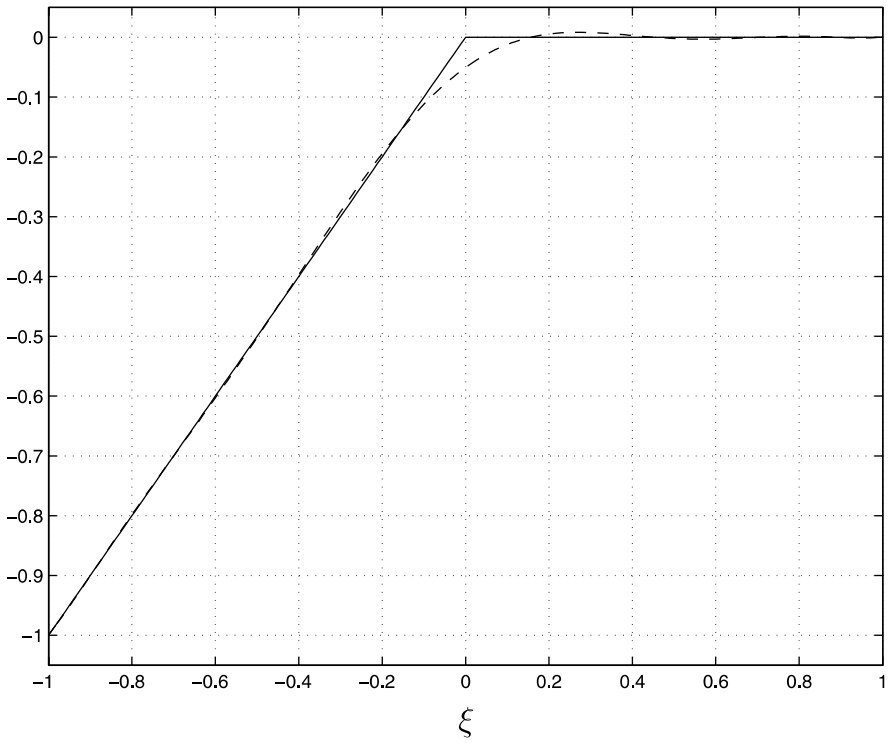


Figure 4. The graphs of $\min\{0, \xi\}$ and $M_{4,1}(\xi)$.

4.2.2. *Discussions*

Each function obtained from the new Equation (11) is a power series in $\mathcal{T}(\delta t)u_k - u_k$. For instance, if $p = \alpha_k = 1$, then

$$u_{k+1} = -0.1353 + u_k + 0.5[\mathcal{T}(\delta t)u_k - u_k] - 0.3827[\mathcal{T}(\delta t)u_k - u_k]^2.$$

Therefore, a considerable fraction of the total computation cost would contribute to expanding the powers of $\mathcal{T}(\delta t)u_k - u_k$. Suppose that u_k has only two terms, i.e., $u_k(x) = c_1 \exp(a_1\|x + x_1\|^2) + c_2 \exp(a_2\|x + x_2\|^2)$. Then expanding $[\mathcal{T}(\delta t)u_k - u_k]^2$ would give another series of Gaussian functions consisting of ten terms, i.e., $\hat{c}_1 \exp(\hat{a}_1\|x + \hat{y}_1\|^2) + \cdots + \hat{c}_{10} \exp(\hat{a}_{10}\|x + \hat{y}_{10}\|^2)$, where $\hat{c}_i$, $\hat{a}_i$ and $\hat{y}_i$ are the ith entries of, respectively, C, A and X:

$$\begin{aligned} C = \Big(& c_1^2 b_1^n, c_1^2, c_2^2 b_2^n, c_2^2, -2c_1^2 b_1^{\frac{n}{2}}, -2c_2^2 b_2^{\frac{n}{2}}, \\ & 2c_1c_2(b_1b_2)^{\frac{n}{2}} e^{\frac{a_1a_2b_1b_2}{a_1b_1+a_2b_2}\|x_1-x_2\|^2}, -2c_1c_2 b_1^{\frac{n}{2}} e^{\frac{a_1a_2b_1}{a_1b_1+a_2}\|x_1-x_2\|^2}, \\ & -2c_1c_2 b_2^{\frac{n}{2}} e^{\frac{a_1a_2b_2}{a_1+a_2b_2}\|x_2-x_2\|^2}, 2c_1c_2 e^{\frac{a_1a_2}{a_1+a_2}\|x_1-x_2\|^2} \Big)^\top, \end{aligned} \tag{12}$$

$$A = (2a_1b_1, 2a_1, 2a_2b_2, 2a_2, a_1b_1 + a_1, a_2b_2 + a_2,$$
$$a_1b_1 + a_2b_2, a_1b_1 + a_2, a_1 + a_2b_2, a_1 + a_2)^\top,$$
$$X = \Big(x_1, x_1, x_2, x_2, x_1, x_2, \frac{a_1b_1x_1 + a_2b_2x_2}{a_1b_1 + a_2b_2}, \frac{a_1b_1x_1 + a_2x_2}{a_1b_1 + a_2},$$
$$\frac{a_1x_1 + a_2b_2x_2}{a_1 + a_2b_2}, \frac{a_1x_1 + a_2x_2}{a_1 + a_2}\Big),$$

with $b_i = (1 - 4a_i\delta t)^{-1}$ for $i = 1, 2$. In computing the above vectors, the cost that grows with n contributes to the vector additions for C and X only. This is rather inexpensive compared with that of the multiplications and evaluating the exponentials.

The expansion of $[\mathcal{T}(\delta t)u_k - u_k]^r$ needs to be computed using r nested loops in a computer subroutine. Suppose that u_k has N terms. If r and N are too large, then the computer may need an unacceptably long time to complete one iteration. However, even if we could wait for the current iteration to complete, the expansion of $[\mathcal{T}(\delta t)u_k - u_k]^r$ will have

$$\binom{r + 2N - 1}{r} = \frac{(r + 2N - 1)(r + 2N - 2)\cdots(r + 1)}{(2N - 1)!}$$

terms, which makes the next iteration more difficult to proceed.

Fortunately many terms in the expansion of $[\mathcal{T}(\delta)u_k - u_k]^r$ can be ignored because of their negligible values. For instance, in (12),

$$C = (0.33, 1, 0.1, 0.49, -1.15, -0.44, -4.07 \times 10^{-7}, 5.05 \times 10^{-9},$$
$$3.91 \times 10^{-9}, -4.67 \times 10^{-17})^\top$$

for $n = 1$, $\delta t = 0.01$, $c_1 = 1$, $c_2 = -0.7$, $a_1 = -50$, $a_2 = -100$, $x_1 = 0$ and $x_2 = 0.9$. Due to their insignificant values, the last four terms of the expansion can be deleted without introducing much error.

The fidelity of u_k to the exact solution $u(k\delta t, \cdot)$ depends on δt and p. The latter determines the degree of M_{p,α_k}. These parameters also have effects on the length of u_k. Indeed the larger p is, the longer u_k is generated because the higher powers of $\mathcal{T}(\delta t)u_k - u_k$ were expanded. Moreover, the smaller δt is, the larger the coefficients are, (see (12) for instance) and hence the fewer terms can be ignored. In other words, a closer approximation results in u_k with more terms. Therefore δt and p need to be chosen carefully to compromise between the fidelity and length of u_k. Because we do not have a useful rule to select the parameters, they are chosen on an ad hoc basis in the experiments considered below.

4.3. EXPERIMENT RESULTS

We have conducted some experiments on one- and two-dimensional problems to demonstrate the feasibility and potential of the method. The results to be presented in Figures 5–7 show that the method is capable of smoothing out initial functions, and also preserving global minima. The method is, however, not yet suitable for practical problems. It would be difficult in the current stage to compare with other existing global optimization methods.

In these experiments, global minima of u_k are usually very close to that of the cost functions. The shift is mainly due to errors incurred by time discretization and approximation of $\min\{0, \cdot\}$. As we observe, the shift can be severe if δt is too large or p is too small.

Furthermore, u_0 and u_k may not have the same optimum values because $M_{p,\alpha_k}(\xi)$ is nonzero for some $\xi \geqslant 0$. Since $\mathcal{T}(\delta t) f(x^*) - f(x^*) \geqslant 0$, where x^* is a global minimum of f, it may happen that $M_{p,\alpha_k}[\mathcal{T}(\delta) f(x^*) - f(x^*)] \neq 0$, and as a result $u_1(x^*) = f(x^*) + M_{p,\alpha_0}[\mathcal{T}(\delta t) f(x^*) - f(x^*)] \neq f(x^*)$. For this reason, the graphs of u_k shown in the figures shift downwards after each iteration. Of course this will not lead to any trouble if the global minimum of f has been sought already. The differences between the optimum values can be reduced by using an approximating polynomial of higher degree, i.e., using a larger p.

The results displayed in Figures 5–7 are typical in the experiments. Shown in the legend boxes, K denotes the number of iterations that have been carried out, and N_0 and N_K denote the length of u_0 and u_K, respectively. We choose $\alpha_k = \alpha$ for $k = 1, 2, \ldots, K$. Moreover, we neglect a term of u_k whenever the magnitude of its coefficient is smaller than the threshold $\epsilon > 0$.

Figure 5(a) and 5(b) show two univariate initial functions (top curves) possessing three and four Gaussian terms, respectively. The smoothing finishes up with smoothed functions (bottom curves) possessing 851 and 790 terms, and both functions have a single minimum being very close to the global minimum of the corresponding initial function.

Figures 6(a) and 7(a) show two bivariate initial functions possessing six terms. As shown in Figures 6(b) and 7(b), the smoothing finishes up with smoothed functions possessing 601 and 249 terms. Both functions have a single minimum being very close to the global minimum of the corresponding initial function, as seen from the contour plots in Figures 6(d) and 7(d).

Remark 4.1. As we mentioned in Section 3.3, smoothed functions usually evolve very slowly at saddle points with positive Laplacian. Surprisingly, as we observe, perturbation by approximation error tends to speed up the evolution at the saddle points. As a result, u_k becomes unimodal

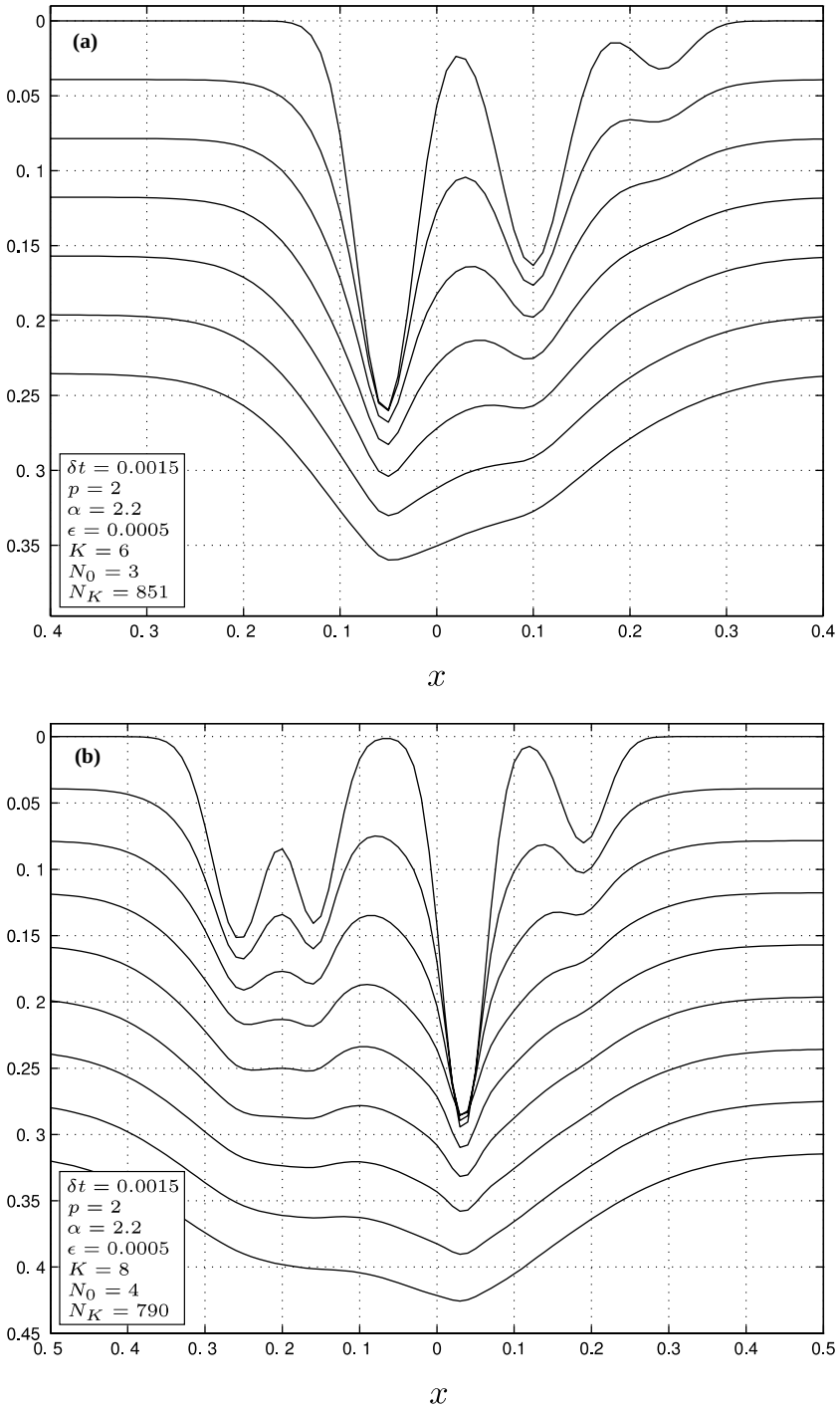
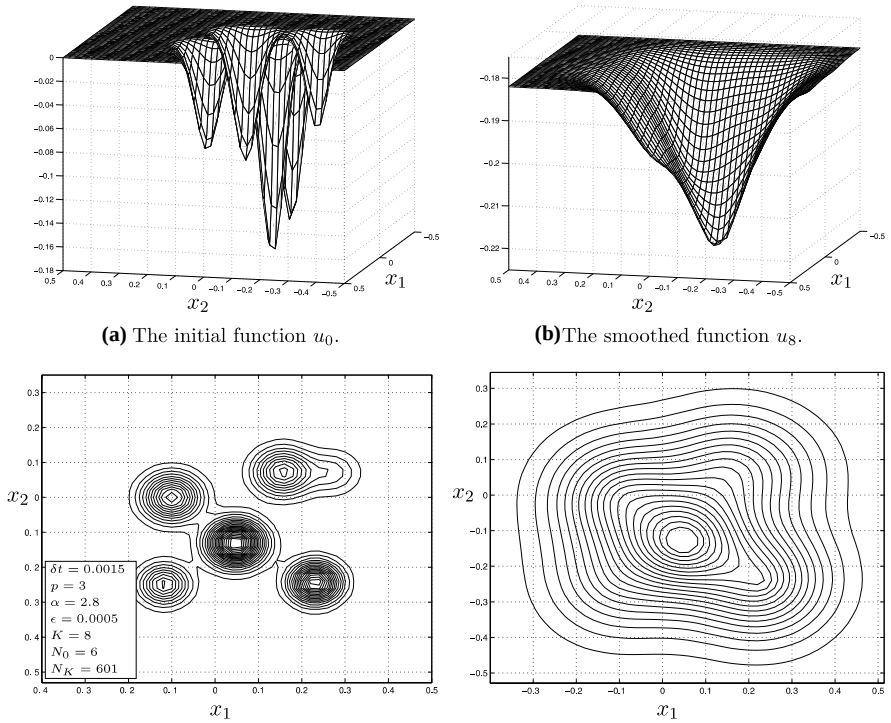


Figure 5. Smoothing of one-variable functions using (11).



(a) The initial function u_0.

(b) The smoothed function u_8.

(c) The contour lines of u_0.

(d) The contour lines of u_8.

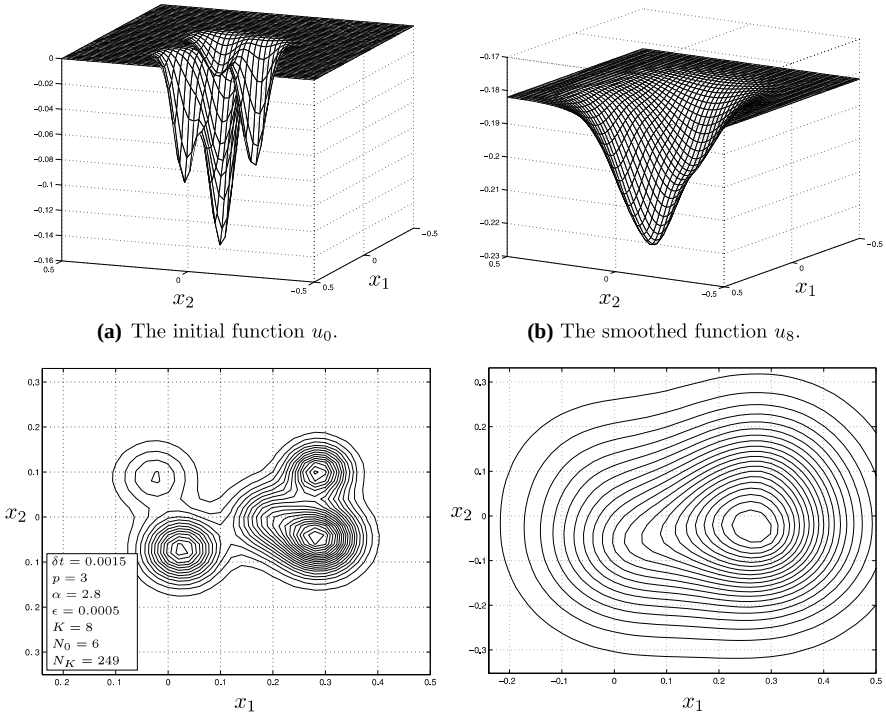
Figure 6. Smoothing of a two-variable function using (11).

after just a few iterations as seen from Figures 6(b) and 7(b). This observation offers a clue of using perturbation to overcome the difficulty.

Remark 4.2. We have carried out some experiments on higher-dimensional problems for n up to 20. As we expect, computation time required for an iteration increases just slightly with n for initial functions having the same number of terms. We plan to access the results in the future. This will probably be achieved by finding global minima of the smoothed and initial functions by exhaustive search, and then making comparisons.

5. Conclusion

A smoothing method of global optimization has been proposed in this paper. The smoothed function of the method is the solution of a heat diffusion equation with external heat source. Using results in theory of viscosity solutions, we have shown that global minima of cost functions always survive through smoothing, i.e., any global minimum of a cost function is again a global minimum of the smoothed function, and both functions



(a) The initial function u_0.

(b) The smoothed function u_8.

(c) The contour lines of u_0.

(d) The contour lines of u_8.

Figure 7. Smoothing of a two-variable function using (11).

have the same optimum values. Several numerical examples have been given to demonstrate this property.

Moreover, we have devised a solution method to compute smoothed functions. The core is an iterative equation, by which smoothed functions can be obtained analytically (without discretization of $\mathbb{R}^n$) for certain classes of initial functions. We have considered initial functions which can be expressed as a series of Gaussian functions. Finally, we have demonstrated the effectiveness and potential of the method with experimental results.

Appendix A. Theory of Viscosity Solutions

In this Appendix, we first give a definition of viscosity solution of initial-value problems with parabolic equations. In Section A.2 we present the proof of our existence and uniqueness result, i.e.. Proposition 3.4. Finally, in Section A.3, we prove two propositions that have been employed in Section 3.

A.1. DEFINITION OF VISCOSITY SOLUTIONS

Let u be a real-valued function defined on $[0, T) \times \mathbb{R}^n$. We consider the viscosity solution of the parabolic PDE

$$\frac{\partial u}{\partial t}(t, x) + F\big[Du(t, x), D^2u(t, x)\big] = 0 \quad \text{for } (t, x) \in (0, T) \times \mathbb{R}^n \tag{13}$$

with the initial condition $u(0, \cdot) = u_0$ on $\mathbb{R}^n$. We recall that Du and D^2u denote, respectively, the gradient vector and Hessian matrix with respect to x; and S^n represents the set of all $n \times n$ symmetric matrices of real numbers. We assume that F is continuous on $\mathbb{R}^n \times S^n$ and **degenerate elliptic**, i.e.,

$$F(p, X + Y) \leqslant F(p, X)$$

for every $X, Y \in S^n$ and $Y \geqslant 0$.

Viscosity solutions of (13) have been widely studied. Interested readers are referred to the papers by Crandall [8,11], Chen [4], Giga [15] and Barles [1]. The first two papers are self-contained expositions of basic theories, concentrating on problems with bounded domains; the last two papers consider problems with unbounded domains.

DEFINITION A.1. A function $u: (0, T) \times \mathbb{R}^n \to \mathbb{R}$ is a **viscosity subsolution (supersolution)** of (13) on $(0, T) \times \mathbb{R}^n$, if u is upper (lower) semicontinuous on $(0, T) \times \mathbb{R}^n$, and for each $\phi \in C^2((0, T) \times \mathbb{R}^n)$ and maximum (minimum) $(\hat{t}, \hat{x}) \in (0, T) \times \mathbb{R}^n$ of $u - \phi$, we have

$$\phi_t(\hat{t}, \hat{x}) + F\big[D\phi(\hat{t}, \hat{x}), D^2\phi(\hat{t}, \hat{x})\big] \leqslant 0.$$
$$\Big(\phi_t(\hat{t}, \hat{x}) + F\big[D\phi(\hat{t}, \hat{x}), D^2\phi(\hat{t}, \hat{x})\big] \geqslant 0.\Big)$$

Moreover, u is a **viscosity solution** of (13) on $(0, T) \times \mathbb{R}^n$ if it is both viscosity subsolution and supersolution of (13) on $(0, T) \times \mathbb{R}^n$.

DEFINITION A.2. Suppose that u is a viscosity subsolution (supersolution) of (13) on $(0, T) \times \mathbb{R}^n$. If u is upper (lower) semicontinuous on $[0, T) \times \mathbb{R}^n$ and satisfies

$$u(0, x) \leqslant u_0(x), \quad \big(u(0, x) \geqslant u_0(x)\big)$$

for every $x \in \mathbb{R}^n$, then it is a **viscosity subsolution (supersolution)** of (13). Moreover, u is a **viscosity solution** of (13) if it is both viscosity subsolution and supersolution of (13).

A clear presentation of the motivations behind these definitions can be found in [13]. Alternate definitions were announced in [10] and [22], which are more appealing in some respects and convenient for certain purposes. It is however not worthwhile to state the alternate definitions here since our work apparently does not enjoy the benefits.

In Definition A.1 a C^2 function ϕ is introduced so that the inequalities depend on derivatives of ϕ only. This does not lead us to define derivatives for u, Therefore u needs not to be differentiable. But it is necessarily continuous on $[0, T) \times \mathbb{R}^n$ since it is both upper and lower semicontinuous on $[0, T) \times \mathbb{R}^n$.

A.2. PROOF OF EXISTENCE AND UNIQUENESS RESULT

To prove the result, we use the following notations : The **upper (lower) semicontinuous envelope** [8,11] of $u: Q \to [-\infty, \infty]$, where $Q \subset \mathbb{R}^n$, is the function defined by

$$u^*(x) = \lim_{r \downarrow 0} \sup \{u(y) : y \in Q, \|y - x\| \leqslant r\}, \tag{14}$$

$$\left(u_*(x) = \lim_{r \downarrow 0} \inf \{u(y) : y \in Q, \|y - x\| \leqslant r\}\right) \tag{15}$$

for each $x \in Q$. The following properties will be used subsequently: 1f v is upper semicontinuous such that $u \leqslant v$ on Q, then $u \leqslant u^* \leqslant v$ on Q. Similarly if v is lower semicontinuous such that $v \leqslant u$ on Q, then $v \leqslant u_* \leqslant u$ on Q.

The existence and uniqueness result will be proved using Proposition 3.3; the latter is a special version of Theorem A.3. The theorem was given in [15] and considers the Equation (13) with F satisfying the following conditions:

(F1) F is continuous on $(\mathbb{R}^n \setminus \{0\}) \times S^n$.
(F2) $F(p, X + Y) \leqslant F(p, X)$ for every $X, Y \in S^n$ with $Y \geqslant 0$.
(F3) $-\infty < F_*(0, 0) = F^*(0, 0) < \infty$ where F_* and F^* are, respectively, the lower and upper semicontinuous envelopes of F, i,e.,

$$F_*(p, X) = \lim_{\epsilon \downarrow 0} \inf \{F(q, Y) : q \neq 0, \|p - q\| \leqslant \epsilon, \|X - Y\| \leqslant \epsilon\}$$

and $F^* = -(-F)_*$. Here $\|X\|$ denotes the operator norm of $X \in S^n$.
(F4) For every $R > 0$,

$$c_R = \sup \{|F(p, X)| : p \neq 0, \|p\| \leqslant R, \|X\| \leqslant R\} < \infty.$$

THEOREM A.3. *Suppose that F satisfies $(F1)$–$(F4)$ Let μ and ω be, respectively, viscosity subsolution and supersolution of* (13) *on* $(0, T) \times \mathbb{R}^n$. *Assume that μ and ω satisfies the conditions:*

(V1) *There exists a number* $K > 0$ *such that* $\mu(t, x) \leqslant K(\|x\| + 1)$ *and* $\omega(t, x) \geqslant -K(\|x\| + 1)$ *for every* $(t, x) \in (0, T) \times \mathbb{R}^n$;

(V2) *There is a modulus* m *such that* $\mu^*(0, x) - \omega_*(0, y) \leqslant m(\|x - y\|)$ *for every* $x, y \in \mathbb{R}^n$; *and*

(V3) *There exists a number* $K > 0$ *such that* $\mu^*(0, x) - \omega_*(0, y) \leqslant K(\|x - y\| + 1)$ *for every* $x, y \in \mathbb{R}^n$.

Then $\mu < \omega$ *on* $(0, T) \times \mathbb{R}^n$.

Now we prove the existence and uniqueness result, i.e., Proposition 3.4. It essentially follows the proofs of Theorem 3.1 of [1] and Theorem 4.1 of [11].

Proof of Proposition 3.4. Consider the initial-value problem associated with the following equation and initial condition:

$$\begin{aligned} &u_t + F_\epsilon(D^2 u) = 0, \quad \text{on } (0, T) \times \mathbb{R}^n, \\ &u = f, \quad \text{in } \{0\} \times \mathbb{R}^n, \end{aligned} \tag{16}$$

where $F_\epsilon(X) = \min\{1/\epsilon, -\min\{0, \operatorname{tr}(X)\}\}$ for $\epsilon > 0$ and $X \in S^n$. It is straight forward to show that $\underline{u}_\epsilon(t, x) = f(x) - t/\epsilon$ and $\overline{u}_\epsilon(t, x) = f(x) + t/\epsilon$ are, respectively, subsolution and supersolution of (16). We now split the proof into three parts.

Part I Here we show that (16) has solution for each $\epsilon > 0$. Define $W(t, x) = \sup\{w(t, x) : \underline{u}_\epsilon \leqslant w \leqslant \overline{u}_\epsilon,\ w \text{ is a subsolution of (16)}\}$ for each $(t, x) \in [0, T) \times \mathbb{R}^n$. The upper semicontinuous envelope W^* is finite because $\underline{u}_\epsilon \leqslant W^* \leqslant \overline{u}_\epsilon$. Hence we see from Proposition 8.2 of [8] that W^* is a subsolution of (16). Because W is the maximal subsolution between $\underline{u}_\epsilon$ and $\overline{u}_\epsilon$, it follows that $W^* \leqslant W$. But $W \leqslant W^*$, therefore $W = W^*$ is a subsolution of (16).

If W_* fails to be a supersolution at $(\hat{t}, \hat{x}) \in (0, T) \times \mathbb{R}^n$, Lemma 9.1 of [8] suggests that for any small $\kappa > 0$ there exists a subsolution W_κ of (16) on $(0, T) \times \mathbb{R}^n$ having the properties

$$\begin{aligned} &W_\kappa \geqslant W \text{ on } \quad (0, T) \times \mathbb{R}^n, \\ &\sup_{(0,T)\times\mathbb{R}^n} (W_\kappa - W) > 0, \\ &W_\kappa(t, x) = W(t, x) \text{ for } (t, x) \in (0, T) \times \mathbb{R}^n, |t - \hat{t}| + \|x - \hat{x}\| \geq \kappa. \end{aligned}$$

Choosing $\kappa \leqslant |\hat{t}|$ small enough and letting $B = \{x \in \mathbb{R}^n : \|x - \hat{x}\| < \kappa\}$, we see that $W_\kappa = W \leqslant \overline{u}_\epsilon$ in $(\{0\} \times B) \cup ([0, T) \times \partial B)$ where ∂B denotes the boundary of B. Theorem A.1 of [4] then suggests that $W_\kappa \leqslant \overline{u}_\epsilon$ on $(0, T) \times B$. Since W is

the maximal subsolution between $\underline{u}_\epsilon$ and $\overline{u}_\epsilon$, we arrive at the contradiction $W_\kappa \leqslant W$ on $(0, T) \times B$. Therefore W_* is a supersolution. The comparison result Theorem A.3 implies that $W^* = W \leqslant W_*$. This shows that $u_\epsilon = W$ is a solution of (16), which is continuous on $[0, T) \times \mathbb{R}^n$.

Part II The solution u_ϵ is locally bounded on $[0, T) \times \mathbb{R}^n$. This is because the functions $\mu, (t, x) = \min_{x \in \mathbb{R}^n} f(x)$ and $\omega(t, x) = f(x)$ are, respectively, subsolution and supersolution of (16). It is apparent that F_ϵ satisfies (F1)–(F4), and μ and ω satisfy (V1)–(V3). Proposition A.3 therefore suggests that, $\mu \leqslant u_\epsilon \leqslant \omega$ on $[0, T) \times \mathbb{R}^n$. This shows our assertion.

Part III Then we can apply the "half-relaxed-limits" method to complete the proof. The method consists in introducing

$$\begin{aligned}
\overline{u}(t, x) &= \limsup_{\epsilon \downarrow 0}{}^* u_\epsilon(t, x) \\
&:= \limsup_{m \downarrow 0} \{u_\epsilon(\tau, \xi) : \epsilon \leqslant m, (\tau, \xi) \in (0, T) \times \mathbb{R}^n, |\tau - t| + \|\xi - x\| \leqslant m\}; \\
\underline{u}(t, x) &= -\limsup_{\epsilon \downarrow 0}{}^* (-u_\epsilon(t, x)).
\end{aligned}$$

In virtue of Part II, they are well-defined with lower bound $f(x^*)$ and upper bound $f(x)$ where x^* denotes a global minimum of f.

Theorem 8.3 of [8] therefore suggests that $\overline{u}$ and $\underline{u}$ are, respectively subsolution and supersolution of the smoothing problem. Note that

$$\begin{aligned}
&\underline{u}(t, x) \leqslant f(x) \leqslant K(\|x\| + 1) \leqslant K(\|x^*\| + 1)(\|x\| + 1); \text{ and} \\
&\overline{u}(t, x) \geqslant f(x^*) \geqslant -K(\|x^*\| + 1) \geqslant -K(\|x^*\| + 1)(\|x\| + 1).
\end{aligned}$$

Proposition 3.3 applies giving the inequality $\overline{u} \leqslant \underline{u}$, Since $\underline{u} \leqslant \overline{u}$ by definitions, this leads us to conclude that $u = \underline{u} = \overline{u}$ is a continuous solution of the smoothing problem. Uniqueness of u is a direct consequence of the comparison result. □

A.3. AUXILIARY RESULTS

In this section we state and prove two propositions which have been employed to prove the global minima preserving property in Section 3.

PROPOSITION A.4. *Let f be the initial function of a smoothing problem satisfying (U1)–(U3), and denote by $\underline{f}$ its convex envelope over $\mathbb{R}^n$. Let $g(t, x) = f(x)$ and $\underline{g}(t, x) = \underline{f}(x)$ for every $(t, x) \in [0, T) \times \mathbb{R}^n$. Then g and $\underline{g}$ are, respectively, viscosity supersolution and subsolution of the problem.*

Proof. Let ϕ be twice continuously differentiable on $(0, T) \times \mathbb{R}^n$, and (t^*, x^*) in $(0, T) \times \mathbb{R}^n$ be a minimum point of $g - \phi$. Since g is constant along the t-axis, we have $\phi_t(t^*, x^*) = g_t(t^*, x^*) = 0$. As a result,

$$\phi_t(t^*, x^*) - \min\left\{0, \operatorname{tr}(D^2\phi(t^*, x^*))\right\} \geqslant 0.$$

Moreover $g(0, x) = f(x)$ for every $x \in \mathbb{R}^n$. Hence g is a viscosity supersolution of the smoothing problem.

Now suppose that (t^*, x^*) is a maximum of $\underline{g} - \phi$. Since $\underline{g}$ is convex on $\mathbb{R}^n$, there exists a (subgradient) vector h such that $\underline{g}(t, x) \geqslant \underline{g}(t^*, x^*) + (x - x^*)^{\mathsf{T}} h$ for every $x \in \mathbb{R}^n$. As a result, we have $\phi(t, x) \geqslant \phi(t^*, x^*) + (x - x^*)^{\mathsf{T}} h$ for all (t, x) sufficiently close to (t^*, x^*). By Taylor expansion of ϕ around (t^*, x^*), this inequality can be written as

$$(x - x^*)^{\mathsf{T}}(p - h) + \tfrac{1}{2}(x - x^*)^{\mathsf{T}} X (x - x^*) + o(\|t - t^*\| + \|x - x^*\|^2) \geqslant 0, \quad (t, x) \to (t^*, x^*),$$

where $p = D\phi(t^*, x^*)$ and $X = D^2\phi(t^*, x^*)$. By sending $(t, x) \to (t^*, x^*)$, we see that $p = h$ and $X \geqslant 0$. Therefore

$$\phi_t(t^*, x^*) - \min\{0, X\} = 0.$$

Moreover $\underline{g}(0, x) = \underline{f}(x) \leqslant f(x)$ for every $x \in \mathbb{R}^n$. Hence $\underline{g}$ is a viscosity subsolution of (3). □

PROPOSITION A.5. *Let u be the viscosity solution of a smoothing problem. Let $\upsilon(t, x) = u(t + \tau, x)$ for $\tau > 0$ and every $(t, x) \in [0, T - \tau) \times \mathbb{R}^n$. If the initial function f satisfies $(U1)$–$(U3)$, then υ is a viscosity subsolution of the smoothing problem on $(0, T - \tau) \times \mathbb{R}^n$.*

Proof. Let $T' = T - \tau$. For every ψ twice continuously differentiable on $(0, T') \times \mathbb{R}^n$ and a maximum $(\hat{t}, \hat{x}) \in (0, T') \times \mathbb{R}^n$ of $\upsilon - \psi$, there exists a maximum $(\tau + \hat{t}, \hat{x})$ of $u - \phi$, where $\phi(t + \tau, x) = \psi(t, x)$. Since u is a subsolution of the smoothing problem, we have

$$\phi_t(\tau + \hat{t}, \hat{x}) \leqslant \min\{0, \Delta\phi(\tau + \hat{t}, \hat{x})\}.$$

Because $\phi_t(\tau + \hat{t}, \hat{x}) = \psi_t(\hat{t}, \hat{x})$ and $\Delta\phi(\tau + \hat{t}, \hat{x}) = \Delta\psi(\hat{t}, \hat{x})$, we conclude that

$$\psi_t(\hat{t}, \hat{x}) \leqslant \min\ \{0, \Delta\psi(\hat{t}, \hat{x})\}.$$

Therefore υ is a subsolution on $(0, T') \times \mathbb{R}^n$. □

Acknowledgements

This research is partially supported by a grant from the Research Grants Council of Hong Kong (Project ID 2050277). The authors would like to thank Professor Alexander G. Ramm for pointing out a mistake in an early version of the present paper.

References

1. Barles, G., Biton, S. and Ley, O. (2003), Uniqueness for parabolic equations without growth condition and applications to the mean curvature flow in $\mathbb{R}^2$, *Journal of Differential Equations*, 187, 456–472.
2. Boyd, J.P. (2001), *Chevyshev and Fourier Spectral Methods*, 2nd edition, Dover.
3. Burden, R.L. and Faires, J.D. (1997), *Numerical Analysis*, 5th edition, Brooks/Cole.
4. Chen, Y.G., Giga, Y. and Goto, S. (1991), Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations, *Journal of Differential Geometry*, 33.
5. Cheney, E.W. (1982), *Introduction to Approximation Theory*. AMS Chelsea Publishing.
6. Cheney, E.W. (1982), *A Course in Approximation Theory*. Brooks/Cole.
7. Coleman, T., Shalloway, D. and Wu, Z. (1994), A parallel build-up algorithm for global energy minimizations of molecular clusters using effective energy simulated annealing, *Journal of Global Optimization*, 4, 171–186.
8. Crandall, M.G. (1997), Viscosity solutions: A primer. In *Viscosity Solutions and Applications*, vol. 1660 of *Lecture Notes in Mathematics*. Springer, Berlin.
9. Crandall, M.G., Evans, L.C. and Lions, P.L. (1983), Viscosity solutions of Hamilton-Jacobi equations, *Transactions of the AMS*, 277(1), 1–42.
10. Crandall, M.G., Evans, L.C. and Lions, P.L. (1984), Some properties of viscosity solutions of Hamilton-Jacobi equations, *Transactions of the AMS*, 282(2), 487–502.
11. Crandall, M.G., Hshii, H. and Lions, P.L. (1992), User's guide to viscosity solutions of second order partial differential equations, *Bulletin* (*New Series*) *of the AMS*, 27(1), 1–67.
12. Curtain, R.F. and Zwart, H.J. (1995), *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer, Berlin.
13. Evans, L.C. (1988), *Partial Differential Equations*, vol. 19 of *Graduate Studies in Mathematics*, American Mathematical Society.
14. Ge, R. and Qin, Y. (1990), The globally convexized filled functions for global optimization. *Applied Mathematics and Computation*, 35, 131–158.
15. Giga, Y., Goto, S., Ishii, H. and Sato, M.H. (1991), Comparison principle and convexity preserving properties for singular degenerate parabolic equations on unbounded domains, *Indiana University Mathematics Journal*, 40(2), 443–469.
16. Horst, R. Pardalos, P.M. and Thoai, N.V. (2000), *Introduction to Global Optimization*. 2nd edition, Kluwer Academic Publishers, Doudrecht.
17. Kevorkian, J. (2000), *Partial Differential Equations: Analytical Solution techniques*, 2nd edition, Springer, Berlin.
18. Kostrowicki, J. and Piela, L. (1991), Diffusion equation method of global minimization: Performance for standard test functions, *Journal of Optimization Theory and Applications*, 69(2), 269–284.
19. Kostrowicki, J. Piela, L. Cherayil, B.J. and Scheraga, H. (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.
20. Kostrowicki, J. and A. Scheraga, H. (1992), Application of the diffusion equation method for global optimization to oligopeptides, *Journal of Physical Chemistry*, 96, 7442–7449.
21. Kostrowicki, J. and Scheraga, H.A. (1996), Some approaches to the multiple-minima problem in protein folding. In Pardalos, P.M. Shalloway, D. and Xue, G. (ed.), *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, American Mathematical Society, pp. 123–132.
22. Lions, P.L. (1983), Optimal control of diffusion processes and hamilton-jacobi-bellman equation, part 1: The dynamic programming principle and applications and part 2: Viscosity solutions and uniqueness, *Communications in Partial Differential Equation*, 8, 1101–1174, 1229–1276.

23. Liu, X. (2001), Finding global minima with a computable filled function, *Journal of Global Optimization*, 19, 151–161.
24. Moré, J.J. and Wu, Z. (1986), Smoothing techniques for macromolecular global optimization. In DiPillo, G. and Gianessi, F. (eds.), *Nonlinear Optimization and Applications*, Plenum Press, Newyork pp. 297–312.
25. Moré, J.J. and Wu, Z. (1997), Global continuation for distance geometry problems, *SIAM Journal on Optimization*, 7, 814–837.
26. Piela, L., Kostrowicki, J. and Scheraga, H. (1989), The multiple-minima problem in the conformational analysis of molecules. Deformation of the potential energy hypersurface by the diffusion equation method, *Journal of Physical Chemistry*, 93, 3339–3346.
27. Shalloway, D. (1992), Packet annealing: A deterministic method for global minimization. Application to molecular conformation. In Floudas, C.A. and Pardalos, P.M. (eds.), *Recent Advances in Global Optimization*, Princeton University Press, NJ. pp. 433–477.
28. Stillinger, F.H. (1985), Role of potential-energy scaling in the low-temperature relaxation behavior of amorphous materials. *Physical Review B*, 32, 3134–3141.
29. Stillinger, F.H. and Weber, T.A. (1988), Nonlinear optimization simplified by hypersurface deformation, *Journal of Statistical Physics*, 52, 1429–1445,
30. Vese, L. (1999), A method to convexify functions via curve evolution, *Communications in Partial Differential Equations*, 24: 1573–1591.
31. Wu, Z. (1996), The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation, *SIAM Journal on Optimization*, 6, 748–768.