<table>
<thead>
<tr>
<th>Title</th>
<th>Smoothing Methods and Their Applications in Numerical Analysis and Optimization: A Survey (Discretization Methods and Numerical Algorithms for Differential Equations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Chen, Xiaojun</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 (2002), 1265: 140-151</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2002-05</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/42082">http://hdl.handle.net/2433/42082</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>

Kyoto University
Smoothing Methods and Their Applications in Numerical Analysis and Optimization: A Survey

Xiaojun Chen

Abstract

This paper presents a brief view of recent applications of smoothing methods in the area of numerical analysis and optimization. We describe various nonsmooth problems and illustrate how to apply smoothing methods to these problems. We summarize properties of smoothing methods which are useful for the convergence analysis and error estimation of smoothing methods.

1 Introduction

In the last decade smoothing methods have been successfully applied to many important problems in the area of numerical analysis and optimization. These problems include

- complementarity problems [7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 21, 29, 32, 38, 39, 47, 49, 51, 52, 56, 62, 63, 68, 71, 72],
- variational inequality problems [6, 30, 31, 44, 55, 61, 64, 76],
- optimal control problems with bound constraints on the control [46, 53],
- nonsmooth Dirichlet problems [23, 26, 28],
- computational fluid dynamics [34],
- shape preserving approximation [35],
- nonsmooth convex programs [8, 24],
- comformal mapping [73],
- semi-infinite programs [70],
- mathematical programs with equilibrium constraints [25, 36, 41, 42, 48],
- the unbounded in optimization [2],
- stochastic programs [5, 22],
- minimizing a sum of Euclidean norms [65], etc.

A common feature shared by these problems is that each problem or its reformulation involves functions which are not differentiable in the sense of Fréchet or Gâteaux [58]. These functions are said to be nonsmooth, and the problems are called nonsmooth problems. Traditional algorithms lack robustness for solving these problems [58].
There has been a growing interest in the study of nonsmooth functions in finite dimensional spaces or functional spaces [27, 28, 33, 43, 45, 50, 59, 63, 69, 75]. A number of numerical methods for solving nonsmooth problems based on the theory of nonsmooth analysis have been developed [1, 18, 19, 20, 29, 37, 40, 57, 60, 62, 74, 77]. Among these methods, smoothing methods approximate the nonsmooth functions by parameterized differentiable functions. By updating the parameters in the numerical methods, many traditional algorithms for smooth problems can be modified to efficiently solve these nonsmooth problems.

This paper presents a brief illustration on how to apply smoothing methods to some typical problems. In Section 2, we describe various problems and smoothing approximations for the problems. In Section 3, we summarize the properties of the smoothing approximations and state some concepts in nonsmooth analysis which are useful for the convergence analysis and error estimation of smoothing methods.

A few words about notations. For two vectors \(x, y \in R^n\), \(x \geq y\) and \(x \geq 0\) denote \(x_i \geq y_i\) and \(x_i \geq 0\) for \(i = 1, 2, \ldots, n\), respectively. Let \(e_i\) be the ith column of the identity matrix \(I \in R^{n \times n}\). Let \(R_+ = \{\epsilon | \epsilon > 0, \epsilon \in R\}\).

## 2 Problems and Smoothing Approximations

In this section, we consider seven important problems in numerical analysis and optimization. We show how to define smoothing functions to approximate these problems.

### 2.1. Complementarity problems

Let \(f : R^n \rightarrow R^n\) be a continuously differentiable function. The complementarity problem is to find a vector \(x\) such that

\[
x \geq 0, \quad f(x) \geq 0 \quad \text{and} \quad x^T f(x) = 0.
\]

This problem is called a nonlinear complementarity problem if \(f\) is a nonlinear function, or a linear complementarity problem if \(f\) is an affine mapping of the form

\[
f(x) = Mx + q,
\]

where \(M \in R^{n \times n}\) and \(q \in R^n\).

There are several ways to formulate the complementarity problems as a system of nonsmooth equations. Among these reformulations, the following two functions \(F\) and \(\tilde{F}\) are well-known, whose components are defined by

\[
F_i(x) = \min(x_i, f_i(x))
\]

and

\[
\tilde{F}_i(x) = \frac{1}{2} \left( x_i + f_i(x) - \sqrt{x_i^2 + (f_i(x))^2} \right).
\]

The two functions have the same growth rate by the following inequalities [71]:

\[
\frac{1}{\sqrt{2} + 2} |F_i(x)| \leq \frac{1}{2} |\tilde{F}_i(x)| \leq \frac{\sqrt{2} + 2}{2} |F_i(x)|.
\]

Each \(F_i\) can be written as

\[
F_i(x) = x_i - \max(0, x_i - f_i(x)),
\]

which is piecewise smooth, whose nondifferentiable points form the set:

\[
\{x | x_i = f_i(x) \text{ and } e_i \neq f'_i(x) \}.
\]

The function \(\tilde{F}_i\) is differentiable everywhere except at the point

\[
\{x | x_i = f_i(x) = 0\}.
\]
Two typical smoothing functions $H$ and $\tilde{H}$ approximating to these two nonsmooth functions have the components
\[
H_i(x, \epsilon) = \frac{1}{2} \left( x_i + f_i(x) - \sqrt{(x_i - f_i(x))^2 + 4\epsilon^2} \right)
\]
and
\[
\tilde{H}_i(x, \epsilon) = \frac{1}{2} \left( x_i + f_i(x) - \sqrt{x_i^2 + (f_i(x))^2 + 2\epsilon^2} \right).
\]
For every $\epsilon > 0$, the functions $H$ and $\tilde{H}$ are continuously differentiable with respect to $x$ in $R^n$. Moreover, for all $x \in R^n$ we have
\[
0 \leq F_i(x) - H_i(x, \epsilon) \leq \epsilon
\]
and
\[
0 \leq \tilde{F}_i(x) - \tilde{H}_i(x, \epsilon) \leq \frac{1}{\sqrt{2}}\epsilon.
\]
The smoothing approximations for complementarity problems can also be applied to the problems which involve complementarity problems. For example, mathematical programs with equilibrium constraints [25, 36, 41, 42, 48].

2.2. Variational inequality problems with box constraints
Let $l \in \{R \cup \{-\infty\}\}^n$ and $u \in \{R \cup \{\infty\}\}^n$ be two vectors which satisfy $l \leq u$ and $l \neq u$. Then
\[
X = \{x \in R^n | l \leq x \leq u\}
\]
is called a box in $R^n$. Let $f : D \subset R^n$ be a continuously differentiable function defined on the open set $D \subset R^n$ containing $X$. This problem is to find a vector $x^* \in X$ such that
\[
(y - x^*)^T f(x^*) \geq 0 \quad \text{for} \quad y \in X.
\] (2)

When $l_i = -\infty, u_i = \infty$, for $i = 1, 2, \ldots, n$, this problem reduces to the system of nonlinear equations
\[
f(x) = 0.
\]
When $l_i = 0, u_i = \infty$ for $i = 1, 2, \ldots, n$, this problem reduces to the nonlinear complementarity problem (1). Moreover, if $f$ is the gradient of a function $\phi : R^n \rightarrow R$, this problem becomes the stationary point problem of the following minimization problem with box constraints:

minimize $\phi(x)$
subject to $x \in X$.

We can define two reformulations of this problem as a system of nonsmooth equations:
\[
F(x) = x - \Pi_X(x - f(x)) = 0
\]
or
\[
\tilde{F}(x) = f(\Pi_X(x - f(x))) + x - f(x) - \Pi_X(x - f(x)) = 0,
\]
where $\Pi_X(x)$ denotes the projection of the vector $x$ onto $X$, which can be written as
\[
(\Pi_X(x))_i = \begin{cases} 
  l_i, & z_i \leq l_i \\
  z_i, & l_i < z_i < u_i, \quad i = 1, 2, \ldots, n,
  u_i, & z_i \geq u_i
\end{cases}
\]
The projection is the only nonsmooth term in $F$ and $\tilde{F}$. Hence we can derive a smoothing function for this problem by smoothing the projection. In particular, we have
\[
H(x, \epsilon) = x - P(x - f(x), \epsilon)
\]
\( \tilde{H}(x, \epsilon) = f(P(x - f(x), \epsilon)) + x - f(x) - P(x - f(x), \epsilon) \)

where

\[
R_i(x, \epsilon) = \frac{1}{2} \left( \sqrt{(l_i - z_i)^2 + 4\epsilon^2} - \sqrt{(u_i - z_i)^2 + 4\epsilon^2} + l_i + u_i \right) \quad i = 1, 2, \ldots, n.
\]

For every \( \epsilon > 0 \), \( P \) is continuously differentiable with respect to \( x \) in \( \mathbb{R}^n \), and so are \( H \) and \( \tilde{H} \). Moreover, there are two positive constants \( c_i \) and \( \tilde{c}_i \), which are only dependent of \( l_i \) and \( u_i \), such that

\[
|F_i(x) - H_i(x, \epsilon)| \leq c_i \epsilon,
\]

and

\[
|\tilde{F}_i(x) - \tilde{H}_i(x, \epsilon)| \leq \tilde{c}_i \epsilon.
\]

### 2.3. Optimal Control Problems with Bound Constraints on the Control

Let \( \Omega \subset \mathbb{R}^m \) be a closed and bounded convex set. Let \( K \) be a completely continuous map from \( L^\infty(\Omega) \) to \( C(\Omega) \), and \( \Phi \) be the map on \( C(\mathbb{R}^m) \) given by

\[
\Phi(K(x))(t) = \begin{cases} 
  l(t), & K(x)(t) \leq l(t) \\
  K(x)(t), & l(t) \leq K(x)(t) \geq u(t) \\
  u(t), & K(x)(t) \geq u(t), 
\end{cases}
\]

for given \( l \) and \( u \) in \( C(\Omega) \). This problem is to find \( x \in C(\Omega : \mathbb{R}^m) \) such that

\[
F(x) = x(t) - \Phi(K(x))(t) = 0, \quad \text{on} \quad \Omega.
\]

A paradigm for problems of the form (3) is the integral equation with

\[
K(x)(t) = \int_{\Omega} k(t, s)x(s)ds,
\]

where \( k \in C(\Omega \times \Omega) \) is a smooth kernel function. Discretization of this problem gives the variational inequality problem (2).

### 2.4. Nonsmooth Dirichlet Problems

Let \( \Omega \) be a bounded domain in \( \mathbb{R}^2 \) with a Lipschitz boundary \( \partial \Omega \). Given a real number \( \lambda \), this problem is to find \( u \) such that

\[
\begin{cases} 
  -\Delta u + \lambda \xi(u) = f(x, y) & \text{in} \ \Omega \\
  u = g(x, y) & \text{on} \ \partial \Omega, 
\end{cases}
\]

where

\[
\xi(u) = \begin{cases} 
  u^p, & u \geq 0 \\
  0, & u < 0 
\end{cases}
\]

and \( p \in (0, 1] \) is a positive number. Such problem is related to reaction-diffusion problems [3, 4] and to MHD(magnetohydrodynamics) equilibria [66]. If \( p = 1 \), the term \( \xi(u) = \max(0, u) \) is Lipschitz continuous. We can define two smoothing approximations for \( \max(0, u) \) as

\[
\phi(u, \epsilon) = \frac{1}{2}(u + \sqrt{u^2 + 4\epsilon^2})
\]

and

\[
\tilde{\phi}(u, \epsilon) = \begin{cases} 
  \frac{1}{2\epsilon}(\frac{\epsilon}{2} + u)^2, & |u| \leq \frac{\epsilon}{2} \\
  \max(0, u), & \text{otherwise}. 
\end{cases}
\]

For \( p \in (0, 1) \), \( \xi \) is not Lipschitz continuous. To use the smoothing approximations \( \phi \) and \( \tilde{\phi} \), a Lipschitz reformulation was introduced in [23]:

\[
\begin{cases} 
  -\Delta u + \lambda \max(0, v) = f(x, y) & \text{in} \ \Omega \\
  u = \psi(v) & \text{in} \ \Omega, \\
  u = g(x, y) & \text{on} \ \partial \Omega 
\end{cases}
\]
\[ \psi(v) = \begin{cases} v^{1/p}, & v \geq 0 \\ 0, & v < 0 \end{cases} \]

which is continuously differentiable.

2.5. Computational Fluid Dynamics

We consider Euler equations for flow through a nozzle of length \( L \). The nozzle is a surface of revolution about the \( x \)-axis with cross section area \( S(x) \).

The governing equations for Quasi-One Dimensional Euler flow in conservative variables are

\[
\frac{\partial (\rho S)}{\partial t} + \frac{\partial (\rho u S)}{\partial x} = 0
\]

\[
\frac{\partial (\rho u S)}{\partial t} + \frac{\partial [(\rho u^2 + p)S]}{\partial x} = p \frac{dS}{dx}
\]

\[
\frac{\partial (\rho ES)}{\partial t} + \frac{\partial (\rho uHS)}{\partial x} = 0
\]

where \( \rho(x,t) \) is density, \( p(x,t) \) is pressure, \( u(x,t) \) is velocity,

\[ E(x,t) = \frac{c^2}{\gamma(\gamma-1)} + \frac{u^2}{2}, \]

is total energy and

\[ H(x,t) = \frac{1}{\rho}(\rho E + p) \]

is stagnation enthalpy. Here \( c = \sqrt{\gamma p/\rho} \) is the speed of sound, and \( \gamma \) is the ratio of the specific heat at constant pressure to the specific heat at constant volume.

In many applications, the cross-sectional area of the flow domain is nonsmooth. For instance, we consider the following example [34]:

\[ S(x) = \begin{cases} 1 + 4(x-1)^2, & 0.5 < x < 1.5 \\ 2, & \text{otherwise} \end{cases} \]

The boundary conditions are supersonic flow in the inlet. It is easy to find that

\[ S(x) = \min(2, 1 + 4(x-1)^2). \]

Since \( 1 + 4(x-1)^2 \) is continuously differentiable, we can define the smoothing functions in the similar way as for complementarity problems. In particular, we have

\[ \tilde{S}(x, \epsilon) = \frac{1}{2} \left( 3 + 4(x-1)^2 - \sqrt{(1-4(x-1)^2)^2 + 4\epsilon^2} \right). \]

Replacing \( S(x) \) by \( \tilde{S}(x, \epsilon) \) in the governing equations, we obtain a smoothing approximation for this problem.

2.6 Shape Preserving Approximation

This problem arises from practical applications in computer aided geometric design where one has not only to approximate data points but also to achieve a desired shape of a curve or surface. A special case of shape preserving approximation is one-dimensional convex best interpolation, which is to find a real valued function that is convex and passes through given points in \( R \). We write this problem as a constrained minimization problem:

\[
\begin{align*}
\text{minimize} & \quad \|f''\|_2 \\
\text{subject to} & \quad f(t_i) = y_i, \ i = 0, 1, \ldots, n + 1 \\
& \quad f \text{ is convex on } [a, b] \\
& \quad f \in W^{2,0}[a, b],
\end{align*}
\]
where \( a = t_0 < t_1 < \cdots < t_{n+1} = b \) and \( y_i, i = 0, 1, \ldots, n + 1 \) are given numbers, \( \| \cdot \|_2 \) is the Lebesgue \( L^2[a, b] \) norm, and \( W^{2,2}[a, b] \) denotes the Sobolev space of functions with absolutely continuous first derivatives and second derivatives in \( L^2[a, b] \), equipped with the norm being the sum of the \( L^2[a, b] \) norms of the function, its first and its second derivatives.

Empolying the normalized B-splines \( B_i \) of order two associated with \( (t_i, y_i), i = 0, 1 \ldots, n+1 \), and the corresponding second divided differences \( d_i \), the problem can be rewritten as nonsmooth equations

\[
F(x) = G(x) - d = 0
\]

where

\[
G_i(x) = \int_a^b \left( \sum_{i=1}^n x_j B_j(t) \right) B_i(t) \, dt.
\]

Here \( (z)_+ \) denotes \( \max(0, z) \).

Using the solution \( x^* \) of (4), we can define the second derivative of the desired function as

\[
f''(t) = \left( \sum_{i=1}^n x_j^* B_j(t) \right)_+.
\]

It is well-known that a function \( f \) is convex if and only if the second derivative \( f'' \) is nonnegative.

The function \( G \) is nonsmooth. To see it, we consider the following example. Let \( t_i = i + 1, i = 0, 1, 2, 3 \). The B-splines are defined by

\[
B_1(t) = \begin{cases} t - 1, & t \in [1, 2] \\ 3 - t, & t \in [2, 3] \end{cases}
\]

and

\[
B_2(t) = \begin{cases} t - 2, & t \in [2, 3] \\ 4 - t, & t \in [3, 4] \end{cases}
\]

In this case, the function \( G \) is given by

\[
G(x) = G(x_1, x_2) = \left( \int_1^2 (x_1 B_1(t) + B_1(t)) \, dt + \int_2^3 (x_1 B_1(t) + x_2 B_2(t)) \, dt \right) + \left( \int_2^3 (x_1 B_1(t) + x_2 B_2(t)) \, dt + \int_3^4 (x_2 B_2(t)) \, dt \right).
\]

The function \( G \) is not differentiable at \( (x_1, x_2) = (0, 0) \). If we replace the term \((\cdot)_+\) by a smoothing approximation as we have done for \( \max(0, u) \), we can get a smoothing function for \( F \).

2.7. Stochastic Programs

A version of two-stage stochastic program with recourse is

\[
\begin{align*}
\text{minimize} & \quad c^T x + F(x) \\
\text{subject to} & \quad Ax = b, \quad x \geq 0
\end{align*}
\]

where

\[
F(x) = \sum_{i=1}^N Q(x, w_i) \rho_i.
\]

Here \( \rho_i \geq 0, \sum_{i=1}^N \rho_i = 1 \) and \( Q \) is calculated by finding for given decision \( x \) and even \( w \), an optimal recourse \( y \in \mathbb{R}^{n_2} \), namely

\[
Q(x, w) = \max \{ (h(w) - T(w))^T z \mid W^T z \leq q \}.
\]

The cost coefficient vector \( c \in \mathbb{R}^n \), the constrained matrix \( A \in \mathbb{R}^{m \times n} \) and the vector \( b \in \mathbb{R}^m \) in the first stage (a master problem), and the associated cost coefficient vector \( q \in \mathbb{R}^{n_2} \) and
the recourse matrix $W \in \mathbb{R}^{m_2 \times n_2}$ in the second stage (a recourse problem) are assumed to be deterministic. In the second stage, the demand vector $h(\cdot) \in \mathbb{R}^{m_2}$ and the technology matrix $T(\cdot) \in \mathbb{R}^{m_2 \times n_2}$ are allowed to depend on the random vector $w \in \Omega \subset \mathbb{R}^\epsilon$.

The function $F$ presents the expected value of minimum extra cost based on the first-stage decision and random events, which is convex and nonsmooth. Two smoothing approximations to $F$ were defined in [5, 22]. One of them is given by [22]

$$
\tilde{Q}(x, w, \epsilon) = \max \{-\frac{\epsilon}{2} z^T z + (h(w) - T(w)x)^T z \mid W^T z \leq q\}
$$

and

$$
H(x, \epsilon) = \sum_{i=1}^{N} \tilde{Q}(x, w_i, \epsilon) \rho_i.
$$

Assume that the feasible set $Z$ of the second stage is bounded. Let

$$
\beta \geq \max_{z \in Z} z^T z.
$$

Then we can show that for every $\epsilon > 0$, $F$ is continuously differentiable and for every $x$, there is an $\bar{\epsilon}(x) > 0$ such that for any $\epsilon \in (0, \bar{\epsilon}(x)]$

$$
H(x, \epsilon) \leq F(x) \leq H(x, \epsilon) + \frac{1}{2} \beta \bar{\epsilon}(x).
$$

3 What are Good Smoothing Approximations

For a nonsmooth function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, a good smoothing function $H : \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}^n$ should have the following three properties.

P1. For every $\epsilon > 0$, $H(\cdot, \epsilon)$ is continuously differentiable with respect to $x \in \mathbb{R}^n$.

P2. There is a constant $c > 0$ such that $\|F(x) - H(x, \epsilon)\| \leq c \epsilon$ for $x \in \mathbb{R}^n$ and $\epsilon \in \mathbb{R}_+$.

P3. For every $x \in \mathbb{R}^n$, the limit

$$
\lim_{\epsilon \downarrow 0} H'(x, \epsilon)
$$

exists, say $F^\alpha(x)$, and satisfies

$$
\lim_{h \rightarrow 0} \frac{F(x + h) - F(x) - F^\alpha(x)h}{h} = 0.
$$

The first property states that $H$ is a smoothing function, the second property implies that the error of $H(x, \epsilon)$ to $F(x)$ is bounded by the smoothing parameter $\epsilon$ and the third property is required for designing locally fast convergent algorithms. These smoothing functions discussed in Section 2 satisfy the three properties.

For a nonsmooth problem, we can construct many smoothing approximations of the nonsmooth functions involved in the problem, and design many smoothing algorithms to solve the problem. Using smoothing approximations satisfying the three properties, we can obtain globally and superlinearly convergent algorithms for solving the nonsmooth problems.

The system of nonsmooth equations

$$
F(x) = 0
$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, provides the prime candidate for illustrating the methodology of smoothing methods. For this reason, let us end this paper by considering how to use smoothing approximations to design smoothing methods for solving nonsmooth equations.

Most smoothing methods for solving nonsmooth equations include three steps:
1. **Newton Step** Find a solution $d^k$ of the system of linear equations

$$F(x^k) + F^o(x^k)d = 0. \quad (5)$$

Use the smoothing parameter $\epsilon_k$ to check whether $\|F(x^k + d^k)\|/\|F(x^k)\|$ is small enough. If it is true, let $x^{k+1} = x^k + \hat{d}^k$, otherwise perform Step 2.

2. **Global Smoothing Step** Find a solution $d^k$ of the system of linear equations

$$F(x^k) + \epsilon_k H'(x^k, \epsilon_k)d = 0. \quad (6)$$

Let $m_k$ be the smallest nonnegative integer $m$ such that

$$\|H(x^k + \rho^md^k, \epsilon_k)\|^2 - \|H(x^k, \epsilon_k)\|^2 \leq -\sigma \rho^m \|F(x^k)\|^2, \quad (7)$$

where $\sigma, \rho \in (0, 1)$. Set $t_k = \rho^{m_k}$ and $x^{k+1} = x^k + t_k d^k$.

3. **Update Smoothing parameter $\epsilon_k$**

If $F$ is continuously differentiable, then (P3) implies that $F^o(x) = F'(x)$. Thus, the Newton step (5) is a generalization of the Newton method. Hence, the algorithm will have fast local convergent rate. Using the smoothing function $H$ in the second step, which satisfies (P1)-(P2), will ensure that the solution $d^k$ of (6) is a descent direction, that is, there exists a finite nonnegative integer $m_k$ such that (7) holds. Updating $\epsilon$ in Step 3 makes connection between the Newton step and the global smoothing step. We can show that the smoothing methods are globally and superlinearly convergent. Hence these methods are not only highly efficient but are also robust.

**References**


