ITERATIVE METHODS FOR SOME LARGE SCALE GENERALIZED EQUATIONS

R.S. Womersley

1. INTRODUCTION

Let T be a set valued mapping (multifunction) of $\mathbb{R}^n$ into $\mathbb{R}^n$ (that is $T(x) \subseteq \mathbb{R}^n$ for all $x \in \mathbb{R}^n$). Consider the problem of finding a zero of the map T, that is a point $x^* \in \mathbb{R}^n$ which satisfies the *generalized equation*

$$(1.1) \qquad 0 \in T(x^*) .$$

Such problems frequently arise as necessary conditions for optimization problems, so the continuity properties of the solution sets are of considerable interest (see [3] for example). However here the interest is in numerical methods for calculating x^*. In particular attention is restricted to those maps T where the problem of finding x^* is equivalent to the problem of minimizing some function F.

Consider the maps T which are the generalized gradient $\partial F(x)$ of a locally Lipschitz function $F:\mathbb{R}^n \to \mathbb{R}$. For any $x \in \mathbb{R}^n$ $\partial F(x)$ is a non-empty compact convex set in $\mathbb{R}^n$, and the mapping ∂F is upper semi-continuous (Clarke [1]). A locally Lipschitz function is differentiable almost everywhere, and $\partial F(x)$ is a singleton (the gradient of F) if and only if F is differentiable at x. Also if F is convex then the generalized gradient is the subdifferential of F. A necessary condition for a point x^* to be a local minimizer of F is that $0 \in \partial F(x^*)$, so solving (1.1) equivalent to finding a stationary point of F.

2. MONOTONICITY AND CONVEXITY PROPERTIES

It is commonly assumed that the map T satisfies various monotonicity properties. These properties can be used to partially characterize the maps T for which there exists a function F with $T = \partial F$, and to guarantee existence and uniqueness of solutions to (1.1). If $T = \partial F$ for some function F then the monotonicity properties are equivalent to convexity properties of F.

A set valued mapping T of $\mathbb{R}^n$ into $\mathbb{R}^n$ is

(i) *monotone* iff $(u-v)^T(x-y) \geq 0 \quad \forall x,y \ , \ \forall u \in T(x) \ , \ \forall v \in T(y)$.

(ii) *strictly monotone* iff $(u-v)^T(x-Y) > 0 \quad \forall x,y \ , \ x \neq y \ ,$
$\forall u \in T(x) \ , \ \forall v \in T(y)$.

(iii) *strongly monotone* iff there exists a non-negative function γ such that $\gamma(\alpha) \to \infty$ as $\alpha \to \infty$, $\gamma(\alpha) = 0$ implies $\alpha = 0$ and

$$(u-v)^T(x-y) \geq \|x-y\|\gamma(\|x-y\|) \quad \forall x,y \quad \forall u \in T(x) \ , \ \forall v \in T(y) \ .$$

(iv) *maximal monotone* if the graph of T is not properly contained in the graph of any other monotone mapping.

(v) *cyclically monotone* if for any $m > 0$

$$(x^{(1)}-x^{(0)})^T u^{(0)} + (x^{(2)}-x^{(1)})^T u^{(1)} + \ldots + (x^{(0)}-x^{(m)})^T u^{(m)} \leq 0$$

for any set of points $x^{(i)}$, $i = 0,1,\ldots,m$ with $u^{(i)} \in T(x^{(i)})$.

Any cyclically monotone map is monotone. However if T is linear, so $T(x) = Ax+b$, then T is monotone if and only if the symmetric part of A, namely $\frac{1}{2}(A+A^T)$, is positive semi-definite, whilst T is cyclically monotone if and only if A is symmetric and positive semi-definite.

The following result from Rockafellar [5] characterizes the maps T which are the subdifferentials of a convex function. To the author's knowledge a corresponding result characterizing the maps T which are the generalized gradient of a locally Lipschitz function is not available.

PROPOSITION 2.1: *A set valued mapping* T *from* $\mathbb{R}^n$ *into* $\mathbb{R}^n$ *is the subdifferential of a convex function* $F:\mathbb{R}^n \to \mathbb{R}$ *if and only if* T *is maximal cyclically monotone, moreover the function* F *is unique up to an additive constant.*

Note that the subdifferential of a convex function $F:\mathbb{R}^n \to \mathbb{R}$ is maximal monotone [5]. Also if F is locally Lipschitz then F is convex if and only if ∂F is monotone [1]. Relationships between the monotonicity of ∂F and the convexity of F are summarized in the following proposition.

PROPOSITION 2.2: *Let* $F:\mathbb{R}^n \to \mathbb{R}$ *be locally Lipschitz. Then*

(i) F *is (strictly) convex if and only if* ∂F *is (strictly) monotone.*

(ii) *If* ∂F *is strongly monotone then* F *is strictly convex and coercive (that is* $F(x) \to \infty$ *as* $\|x\| \to \infty$ *).*

When F is convex the problem of finding a point x^* with $0 \in \partial F(x^*)$ is equivalent to finding a minimizer of F.

PROPOSITION 2.3: *Let* $F:\mathbb{R}^n \to \mathbb{R}$ *be convex. The following statements are equivalent*

(i) x^* *is a minimizer of* F $(F(y) \geq F(x^*) \quad \forall y \in \mathbb{R}^n)$.

(ii) $0 \in \partial F(x^*)$ $(\exists u \in \partial F(x^*)$ *such that* $u^T(y-x^*) \geq 0 \quad \forall y \in \mathbb{R}^n)$.

(iii) $v^T(y-x^*) \geq 0 \quad \forall v \in \partial F(y), \quad \forall y \in \mathbb{R}^n$.

One also has the following existence and uniqueness results as a direct consequence of proposition 2.2.

PROPOSITION 2.4: *Let* $F:\mathbb{R}^n \to \mathbb{R}$ *be locally Lipschitz. Then*

(i) *if* ∂F *is strongly monotone there exists a unique point* x^* *with* $0 \in \partial F(x^*)$.

(ii) *if* ∂F *is strictly monotone and if a point* x^* *with* $0 \in \partial F(x^*)$ *exists then* x^* *is unique.*

(iii) *if* ∂F *is monotone then the set of solution points satisfying* $0 \in \partial F(x)$ *is a convex set (if it is non-empty).*

A simple example of a function F for which ∂F is strictly monotone but not strongly monotone is $F(x) = e^x$ for $x \in \mathbb{R}$.

3. SPECIAL STRUCTURE

It is often convenient to decompose a nonsmooth function $F:\mathbb{R}^n \to \mathbb{R}$ as

(3.1) $$F(x) = f(x) + Q(x) \quad \forall x \in \mathbb{R}^n,$$

where $f:\mathbb{R}^n \to \mathbb{R}$ is smooth (continuously differentiable) and $Q:\mathbb{R}^n \to \mathbb{R}$ is nonsmooth. Thus

(3.2) $$\partial F(x) = \nabla f(x) + \partial Q(x) .$$

As $\partial Q(x)$ is multivalued if and only if Q is nonsmooth this corresponds to isolating the multivalued part of ∂F in ∂Q.

A multifunction T from $\mathbb{R}^n$ into $\mathbb{R}^n$ is called *polyhedral* if its graph is the union of a finite collection of polyhedral convex sets. If ∂F is decomposed as in (3.2) ∂F is polyhedral if and only if $\nabla f(x) = Ax+b$ and ∂Q is polyhedral. The function Q is *separable* if

$$Q(x) = \sum_{i=1}^{n} q_i(x_i) .$$

If Q is also convex then

$$\partial Q(x) = \sum_{i=1}^{n} \partial q_i(x_i) .$$

Obviously ∂Q is polyhedral if and only if ∂q_i is polyhedral for $i = 1,\ldots,n$.

A particular example of interest is

$$T(x) = \tau Ax + b + \sum_{i=1}^{n} E(x_i) \ , \tag{3.3}$$

where A is symmetric positive definite, and

$$E(\xi) = \begin{cases} \eta_2(\xi - u_0) + H_0 & \xi > u_0 \\ [0, H_0] & \xi = u_0 \\ \eta_1(\xi - u_0) & \xi < u_0 \ , \end{cases} \tag{3.4}$$

where τ , η_1 , η_2 and H_0 are positive constants and u_0 is a known constant. This is a problem where T is polyhedral and strongly monotone and Q is separable. Obviously (3.3) is the subdifferential of the convex function

$$F(x) = \frac{\tau}{2} x^T Ax + b^T x + \sum_{i=1}^{n} q(x_i)$$

where

$$q(\xi) = \begin{cases} \tfrac{1}{2}\eta_2(\xi - u_0)^2 + H_0(\xi - u_0) & \xi \geq u_0 \\ \tfrac{1}{2}\eta_1(\xi - u_0)^2 & \xi \leq u_0 \ . \end{cases}$$

This particular problem arises in an implicit discretization of the enthalpy formulation of the two-phase Stefan problem (see [2] or [4] for example). The parameter τ corresponds to the time step used in the discretization of the problem. The matrix A corresponds to the standard discretization of the Laplacian in 2 or 3 dimensions, so methods for finding a zero of (3.3) must be suitable to large scale problems. As (3.3) is closely related to a linear system it is expected that methods for finding a zero of (3.3) are going to be closely related to methods for solving large sparse systems of linear equations.

4. ALTERNATING VARIABLE METHODS

The classical Gauss-Seidel and successive overrelaxation (SOR) methods for linear systems (see [6] for example) correspond to methods in which F is minimized in each of the coordinate directions in turn. These methods consist of an iterative scheme of the following form.

(i) Let $x^{(0)}$ be given. Set $k = 0$.

(ii) For $i = 1,\dots,n$

Let $y_i^{(k)}$ be the minimizer of

$$F^{(k+1,i)}(y) = F(x_1^{(k+1)},\dots,x_{i-1}^{(k+1)},y,x_{i+1}^{(k)},\dots,x_n^{(k)}) .$$

Set $x_i^{(k+1)} = x_i^{(k)} + w_i^{(k)}(y_i^{(k)} - x_i^{(k)})$.

(iii) If a convergence test is satisfied stop, otherwise set $k = k+1$ and go to step (ii).

The Gauss-Seidel method corresponds to $w_i^{(k)} \equiv 1$ so $x_i^{(k+1)} = y_i^{(k)}$. An SOR method suggested by Elliott [2] takes

$$w_i^{(k)} = \begin{cases} 1 & \text{if } \partial F^{(k+1,i)}(y) \text{ is multivalued for any } y \in [x_i^{(k)},(1-w)x_i^{(k)}+wy_i^{(k)}]. \\ w & \text{otherwise ,} \end{cases}$$

where $w \in (0,2)$. Note that $y_i^{(k)}$ is characterized by $0 \in \partial F^{(k+1,i)}(y_i^{(k)})$.

If ∂F is decomposed as in (3.1) then one can prove convergence of the Gauss-Seidel method if ∇f is strongly monotone, Q is separable and ∂Q is monotone. For the system defined by (3.3) and (3.4) Elliott [2] has proved convergence of his SOR method for any $w \in (0,2)$. The necessity of Q being separable is illustrated by the following example. Let $n = 2$, $b = 0$, $A = 2I$ and

$$\partial Q(x) = \begin{cases} (0,0)^T & \text{if } x_1 + x_2 > 1 \\ \lambda(-2,-2)^T & \text{where } 0 \le \lambda \le 1 \text{ if } x_1 + x_2 = 1 \\ (-2,-2)^T & \text{if } x_1 + x_2 < 1 . \end{cases}$$

Then $\partial F(x) = Ax + \partial Q(x)$ is the generalized gradient of

$$F(x) = \max \{x_1^2 + x_2^2,\ (x_1 - 1)^2 + (x_2 - 1)^2\} ,$$

which has the unique minimizer $x^* = (\frac{1}{2},\frac{1}{2})^T$. However any method which minimizes along the coordinate directions in turn converges to $\bar{x} = (1,0)^T$ from any starting point $x^{(0)} = (\alpha,0)^T$, where $\alpha \in \mathbb{R}$.

Some results on the asymptotic rate of convergence can be obtained by looking at the spectral radius of the projection of the iteration matrix on to the subspace orthogonal to the surface of nondifferentiability of F at the solution (see [2] for example). This requires that a strict complementarity condition hold at the solution, namely $0 \in \text{rel int } \partial F(x^*)$. As the surface of nondifferentiability at the solution is not known a priori it is difficult to use these results to choose the relaxation parameter w in an SOR method.

5. CONJUGATE GRADIENT METHODS

If F is smooth then the conjugate gradient method generates a search direction $s^{(k)}$ at the point $x^{(k)}$ by $s^{(k)} = -g^{(k)} + \beta^{(k)} s^{(k-1)}$, where $g^{(k)} = \nabla F(x^{(k)})$, and $\beta^{(k)} = g^{(k)T} g^{(k)} / g^{(k-1)T} g^{(k-1)}$ with $\beta^{(0)} = 0$. The point $x^{(k+1)}$ is chosen by a line search giving $x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}$, where $\alpha^{(k)}$ minimizes $F(x^{(k)} + \alpha s^{(k)})$.

When F is nonsmooth, but the generalized gradient is known, a "gradient" can be defined by

$$g(x) = \text{Nr } \partial F(x) ,$$

where $\text{Nr}S$ denotes the point of S nearest to the origin in the Euclidean norm. As $\partial F(x)$ is compact and convex $g(x)$ exists and is unique. Now suppose the usual one-sided directional derivative $F'(x;s)$ always exists then [7]

$$\min_{s:\|s\|\le 1} F'(x;s) = \min_{s:\|s\|\le 1} \max_{u\in\partial F(x)} u^T s = -|\text{Nr } \partial F(x)| .$$

Thus $-g(x)$ is the direction of steepest descent at x, echoing the behaviour of the gradient of a smooth function. The definition of $g(x)$ is also natural as one is trying to solve the generalized equation $0 \in \partial F(x)$.

For some problems, in particular that defined by (3.3) and (3.4), one can write down a concise expression for $g(x)$. One can then develop a conjugate gradient method having many of the features of conjugate gradient methods for smooth problems (for example finite termination if F is a piecewise quadratic function and the surfaces where F is nonsmooth are linear) [8]. The conjugate gradient methods have the advantage that they do not require Q to be separable, or a relaxation parameter to be estimated. However the great simplicity of the alternating variable methods is lost. It should be noted that, although conceptually simple, the line search is often the most computationally expensive, so care must be taken to use an efficient implementation. Further details and a comparison of alternating variable methods and conjugate gradient methods for the problem defined by (3.3) and (3.4) can be found in [8].

REFERENCES

[1] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York and Chichester, 1983.

[2] C.M. Elliott, 'On the finite element approximation of an elliptic variational inequality arising from an implicit time discretization of the Stefan problem', *I.M.A. Journal of Numerical Analysis* 1 (1981), 115-125.

[3] P. Huard (ed.), 'Point-to-set maps and mathematical programming', *Mathematical Programming Study* 10 (1979),

[4] A.D. Miller, 'Numerical modelling of solidification', in S.-Å. Gustafson and R.S. Womersley (eds.), *Proceedings of the Centre for Mathematical Analysis*, Vol. 6 , Australian National University, 1984.

[5] R.T. Rockafellar, 'Characterization of the subdifferentials of convex functions', *Pacific Journal of Mathematics* 17 (1966), 497-510.

[6] R.S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.

[7] P. Wolfe, 'A method of conjugate subgradients for minimizing nondifferentiable functions', *Mathematical Programming Study* 3 (1975), 145-173.

[8] R.S. Womersley, 'Iterative methods for a large scale piecewise quadratic optimization problem arising from an implicit time discretization of the Stefan problem', (to appear).

Mathematical Sciences Research Centre
Australian National University
P.O. Box 4, Canberra, A.C.T.
Australia