Numerical Analysis Project
Manuscript NA-81-14

September 2, 1981

AN EFFICIENT ALGORITHM FOR BIFURCATION PROBLEMS OF VARIATIONAL INEQUALITIES

by

H.D. Mittelmann

Numerical Analysis Project
Computer Science Department
Stanford University
Stanford, California 94305

AN EFFICIENT ALGORITHM FOR BIFURCATION PROBLEMS OF VARIATIONAL INEQUALITIES

H. D. Mittelmann*
Abtcilung Mathcmatik
Univcrsität Dortmund
Postfach SO 05 00
D-4600 Dortmund
Fed. Rep. of Gcrmany

Abstract. For a class of variational inequalities on a Hilbcrt space H bifurcating solutions exist and may be characterized as critical points of a functional with respect to the intersection of the level surfaces of another functional and a closed convex subset K of H. In a recent paper [13] we have used a gradient-projection type algorithm to obtain the solutions for discretizations of the variational inequalities. A related but Newton-based method is given here. Global and asymptotically quadratic convergence is proved. Numerical results show that it may be used very efficiently in following the bifurcating branches and that it compares favorably with several other algorithms. The method is also attractive for a class of nonlinear eigenvalue problems ($K = H$) for which it reduces to a generalized Rayleigh-quotient iteration. So some results are included for the path following in turning point problems.

Subject Classifications: AMS (MOS): 65 K 10, 65 N 30, 65 N 25, 65 L 15, 73 H 05; CR: 5.17

This work was supported in part by the Deutsche Forschungsgemeinschaft and by the Department of Energy contract no. DOE-AS03-76-SF00326-PA # 30, while the author was a visitor' at the Computer Science Depaitmcnt of Stanford University.

1. Introduction

In the following we are concerned with the numerical computation of critical points of a functional $f: H \to \mathbb{R}$, H a real Hilbert space, with respect to the intersection of a closed convex set $K \subset H$ and the level surfaces

$$(1.1) \qquad \partial S_\rho = \{u \in H,\ g(u) = \tfrac{1}{2}\rho^2\}$$

of another (even) functional g. For theoreticsl results concerning existence, characterization of critical points, and relations to bifurcation theory we refer to the literature (see, for example, [1, 15, 19]). Under suitable assumptions a critical point u_0 satisfies the variational inequality

$$(1.2) \qquad \lambda_0(\nabla g(u_0), u - u_0) \geq (\nabla f(u_0), u - u_0),\ \forall u \in K,\ \lambda_0 \in \mathbb{R}$$

and we are interested in the case $\lambda_0 > 0$.

Instead of treating the most general case we describe a class of such problems which is important in physical and mechanical applications. Some examples of this type will be considered. later. H will denote a function space of functions u defined on a domain $\Omega \subset \mathbb{R}^N$, $N \geq 1$, and is usually a Sobolev space $H_0^m(\Omega)$ where only for simplicity the zero boundary conditions are included. The set K will be either the whole space or a subset of the form

$$(1.3) \qquad K = \{u \in H,\ u \geq 0 \text{ a.e. in } C,\ u \leq 0 \text{ a.e. in } D\}$$

where $C, D \subset \Omega$, so that K is in fact a closed convex cone with vertex 0.

While in the case $K = H$ several algorithms have been proposed for the determination of the critical points (see, for example, [9] and the papers cited there) and a vast literature deals with the corresponding differential equation problem, the theory for the case $K \neq H$ has only been developed recently (cf. [12] and the references in [13]). A numerical algorithm was given in [13]. Since this method as well as the algorithm to be presented below attack the discretized problem and have no simple analogue in the continuous case we shall restrict ourselves to finite-dimensional Hilbert spaces.

In [13] the problem of computing bifurcating solutions of variational' inequalities was reduced to a standard optimization problem. A simple gradient-projection type method was used for its numerical solution. In section 4 we describe a Newton type method for the general problem considered in [13] and show in section 6 that it may be used very efficiently for following the bifurcation branches for variational inequalities. Since the method is also attractive for the solution of a class of nonlinear eigenvalue problems, we formulate the method for the case $K=\boldsymbol{H}$ first in the next section and present some numerical results in section 5.

The contents of the following sections are

2. An algorithm for variational equalities
3. Convergence proof
4. **An** algorithm for variational inequalities
5. Path following in turning point problems
6. Path following in bifurcation problems for variational inequalities

2. An Algorithm for Variational Equalities

As indicated in the introduction from now on we shall assume that the functionals fand g are either defined on a finite-dimensional Hilbert space $\boldsymbol{H}$ or a problem of the class described above is discretized by, for example, a finite-difference or a finite element method yielding functionals f_h, g_h defined on a space H_h, where $\boldsymbol{h}$ denotes the discretization parameter. We shall assume that H_h may be identified with Euclidean n-space and we shall omit the subscript $\boldsymbol{h}$.

In this and the following section we treat the case $K=\boldsymbol{H}$ in which inequality (1.2) reduces to the variational equality

$$\lambda_0(\nabla g(u_0),u) = (\nabla f(u_0),u), \qquad \forall u \in H. \tag{2.1}$$

The original problem is the determination of critical points u_0 of the functional f with respect to level sets (1.1) of the functional g.

We make now a few general assumptions on f, g and we refer to the last sections where the examples show that the resulting class of problems covers interesting applications. Let the functional f be twice Frechet differentiable on $\boldsymbol{H}$ and let g be of the form

$$g(x) = \tfrac{1}{2}(Bx,x), \qquad x \in H, \tag{2.2a}$$

where $B: H \rightarrow H$ is a linear, symmetric and positive definite operator. The elements of the finite-dimensional space H will henceforth be denoted by x, y etc..

Let there exist a constant $M = M(\rho) > 0$ such that

$$(2.2b) \qquad 0 < (\nabla f(x+y) - \nabla f(x), y) \leq M\|y\|^2, \quad \forall x \in S_\rho, \quad \forall y \in S_{2\rho}, \quad y \neq 0,$$

and for simplicity let M be chosen such that also the following inequality holds

$$(2.2c) \qquad (\nabla f(y), y) \leq M\|y\|^2, \quad \forall y \in \partial S_\rho,$$

here we have used the notation $S_\rho = \{x \in H, g(x) \leq \frac{1}{2}\rho^2\}$. The norms used here and in the following are the Euclidean norm for $x \in H$ and the spectral norm for matrices $A \in L(H)$.

If (2.2a),(2.2b) and

$$(2.3) \qquad f(0) = 0, \qquad \nabla f(0) = 0$$

are satisfied then (2.1) always has the trivial solution and it is well-known that branches of solutions exist bifurcating from the eigenvalues of the linearized problem (cf. [12] and the references in [13]).

We now present an algorithm for the determination of local maxima off on the level surfaces (1.1) which is well defined under the assumptions of Theorem 2.10 below.

The algorithm for variational equalities

Let $x_1 \in \partial S_\rho$, $\rho > 0$, be arbitrary.

1. For $k = 1, 2, \ldots$ compute

$$(2.4a) \qquad p_k = -H_k r_k,$$

where (formally) H_k is the $n \times n$ principal submatrix of the inverse of

$$(2.4b) \qquad D_k = \begin{bmatrix} F_k - \lambda_k B & -Bx_k \\ -x_k^T B & 0 \end{bmatrix}$$

and we have used the notation $r_k=\nabla f(x_k)$, $F_k=\nabla^2 f(x_k)$, $\lambda_k=r_k{}^T x_k/\rho^2$.

2. Determine a steplength $\alpha_k=2^{-j}$ where

(2.4c) $$j=\min\{i\in\mathrm{N}\cup\{0\},\ f(x_k+2^{-i}p_k)-f(x_k)\geq 2^{-i-2}p_k{}^T r_k\}.$$

3. Set

(2.4d) $$x_{k+1}=\rho(x_k+\alpha_k p_k)/\|x_k+\alpha_k p_k\|_B$$

where $\|\ .\ \|_B=(\ ,\ )_B^{1/2}$ and $(\ ,\ )_B$ denotes the scalar product induced by B.

Remark 2.5 Algorithm (2.4) consists of a damped Newton step for the solution of the Kuhn-Tucker equations

(2.6) $$\nabla f(x)\ -\ \lambda Bx\ =\ 0, \qquad -\tfrac{1}{2}x^T Bx\ +\ \rho^2/2\ =\ 0\,,$$

for updating x_k starting from $x=x_k$, $\lambda=\lambda_k$ and a subsequent normalization to return to the level surface ∂S_ρ. The Lagrange multiplier is updated by $\lambda_{k+1}=r_{k+1}{}^T x_{k+1}/\rho^2$. Hence our method corresponds to the inverse iteration method with Rayleigh-quotient shift, while the Picard iteration considered in [6] corresponds to simple inverse iteration. For the matrix eigenvalue problem, i.e. $f(x)=\tfrac{1}{2}(Ax,x)$, A symmetric, it is well known that the latter process exhibits linear convergence ([22] p. 619) while the first possesses locally cubic convergence properties ([22] p. 636, see also [18]). In the generalization to the nonlinear case considered here and in [6], the order stays the same for the ordinary inverse iteration while algorithm (2.4) will be shown to be quadratically convergent.

Remark 2.6 In order to show how a continuous analog of algorithm (2.4) would look, we derive it for the class of problems from [6]:

(2.7a) $$\lambda L(u)\ =\ \varphi(x,u(x)),\quad x\in\Omega, \qquad u(x)=0,\quad x\in\partial\Omega,$$

where $L(u)=-\ \partial_i(a_{ik}(x)\partial_k u(x))\ +\ a(x)u(x)$ with suitable assumptions on a, a_{ik}, f (and using summation over repeated indices in the definition of L). We add the normalization

(2.7b) $$(u,u)\ :=\ \langle L(u),u\rangle\ =\ \rho^2,$$

where $<\ ,\ >$ denotes the L^2-scalar product on Ω. If now, for simplicity, we consider only the undamped case, a function u_k satisfying (2.7b) would be replaced by

$$u_{k+1} = \rho(u_k+p_k)/\|u_k+p_k\|, \quad \|\cdot\| = (\ ,\ )^{1/2},$$

where p_k is the first component of the solution $v=(v_1,v_2)$ of

$$(\varphi_u(x,u_k(x))-\lambda_k L)v_1(x)-v_2L(u_k(x))=-\varphi(x,u_k(x)),\ x\in\Omega,\quad v_1(x)=0,\ x\in\partial\Omega,$$

$$(u_k,v_1)=0,\ \lambda_k=\langle\varphi(\,.\,,u_k),u_k\rangle/\rho^2,$$

which may be obtained by determining y_k, z_k from the two boundary value problems

$$(\varphi_u(x,u_k)-\lambda_k L)y_k(x)=-L(u_k(x)),\quad x\in\Omega,\quad y_k(x)=0,\quad x\in\partial\Omega, \tag{2.8a}$$

$$(\varphi_u(x,u_k)-\lambda_k L)z_k(x)=-\varphi(x,u_k(x)),\quad x\in\Omega,\quad z_k(x)=0,\quad x\in\partial\Omega \tag{2.8b}$$

and then setting

$$p_k \;=\; z_k-(u_k,z_k)(u_k,y_k)^{-1}y_k. \tag{2.9}$$

We observe, however, that the operator on the left-hand side of (2.8) becomes singular at a turning point and that equation (2.8a) cannot be satisfied there. Hence we are treating problem (2.7a) with a special form of the normalization as used in [8]. Then we apply a Newton step, however only for updating u_k. The normalization (2.7b) is responsible for some simplifications in (2.8), (2.9) compared with other choices.

We now state a local convergence theorem for algorithm (2.4). By $\{x\}^{\perp}$ we denote the orthogonal complement of $x\in H$ with respect to the scalar product $(\ ,\ )_B$.

Theorem ***2.10. Let the assumptions** (2.2) **be satisfied for problem*** (2.1) ***and assume that** x_0 **is a solution of*** (2.1) ***for the parameter** λ_0 **and that** $F(x_0)-\lambda_0B$ **is negative definite on** $\{x_0\}^{\perp}$. **For** x_1 **sufficiently close to** x_0 **the sequence** $\{x_k\}$, $k=1,2,\ldots$ **generated by (2.4) converges to** x_0 **and, if** $f\in C^3(U(x_0))$, **then the asymptotic** (Q-)**order of convergence is two.***

Remark 2.11 We have formulated this local theorem for the unrestricted case since the numerical applications we will treat in sections 5 and 6 essentially need only this result. The theorem will be proved in the next section.

3. Convergence Proof

For the proof of Theorem 2.10 we need the following lemma. It suffices to prove it in the case $\rho=1$.

Lemma 3.1. *Under the assumptions of Theorem* 2.10 *let* $U(x_0)$ *be a neighborhood of* x_0 *such that for all* $x\in U(x_0)$ *and* $\lambda(x) = \nabla f(x)^T x/\rho^2$

$$y^T(F(x)-\lambda(x)B)y \leq -\beta\|y\|^2,\ \beta>0,\ \forall y\in\{x\}^{\perp}. \tag{3.2}$$

Let $x_1\in U(x_0)$ *be chosen such that* $\{x\in\partial S_\rho,\ f(x)\geq f(x_1)\}\subset U(x_0)$. *If* $x_k\in\partial S_\rho$ *is generated by algorithm (2.4) with*

$$\alpha_k = \beta/(2M\mathrm{cond}(B)), \tag{3.33}$$

where $\mathrm{cond}(B)=\|B\|\,\|B^{-1}\|$, *then* $x_{k+1}\in\partial S_\rho$ *and*

$$f(x_{k+1})-f(x_k) \geq c\|p_k\|^2,\ c>0. \tag{3.4}$$

Proof: Consider the case $\rho=1$. $x_{k+1}\in\partial S_\rho$ is valid by construction of the algorithm. The following analysis is similar to that in the proof of Lemma 4.1 in [13] and is therefore given in concise form.

For a suitable $\tau\in(0,1)$ we have from (2.2b)

$$f(x_{k+1})-f(x_k)=-\tau^{-1}(-\nabla f(x_k+\tau(x_{k+1}-x_k))+\nabla f(x_k),\tau(x_{k+1}-x_k))+(\nabla f(x_k),x_{k+1}-x_k)$$

$$\geq -M\|x_{k+1}-x_k\|^2 + (\nabla f(x_k),x_{k+1}-x_k) \tag{3.5}$$

$$\geq 2M\|B^{-1}\|(x_k+y_k,x_{k+1}-x_k)_B,\quad y_k=B^{-1}r_k/(2M\|B^{-1}\|).$$

Hence

$$f(x_{k+1})-f(x_k) \geq d_k(x_k+y_k,x_k+\alpha_k p_k-\|x_k+\alpha_k p_k\|_B x_k)_B$$

where $d_k=2M\|B^{-1}\|/\|x_k+\alpha_k p_k\|_B>0$ and we show next that the second term on' the right-hand side is nonnegative. This condition may be rewritten as

$$(3.6) \qquad (1-\|x_k+\alpha_k p_k\|_B)(1+(y_k,x_k)_B) \;+\; \alpha_k(y_k,p_k)_B \;\geq\; 0.$$

Writing the inverse of D_k in (2.4b) as

$$(3.7) \qquad D_k^{-1} \;=\; \begin{bmatrix} H_k & b_k \\ b_k^T & q_k \end{bmatrix}$$

wc deduce that

$$(3.8a) \qquad H_k(F_k-\lambda_k B) \;-\; b_k x_k^T B \;=\; E_n,$$

$$(3.8b) \qquad H_k B x_k \;=\; 0,$$

where E_n is the identity matrix on $\mathbb{R}^n$. Hence $(x_k,p_k)_B=0$ and for $y=-H_k z$ we have $(y,x_k)_B=0$ and from (3.2), (3.8a) $\|y\|^2\leq\beta^{-1}(y,z)$. Applying this result for $z=r_k$ we derive

$$(3.9) \qquad \|p_k\|^2 \;\leq\; \beta^{-1}(p_k,r_k)$$

and (3.8b) gives

$$(3.10) \qquad \|x_k+\alpha_k p_k\|_B^2 \;\leq\; 1+\alpha_k^2\|B\|\,\|p_k\|^2.$$

From (2.2c) we conclude that $(y_k,x_k)_B\leq½$. Hence

$$(1+(y_k,x_k)_B)^2 \leq 2(1+(y_k,x_k)_B) \;+\; \|p_k\|^2\beta^2/(4M^2\|B^{-1}\|^2\|B\|)$$

the last term being nonnegative and thus

$$(1+\alpha_k^2\|p_k\|^2\|B\|)(1+(y_k,x_k)_B)^2 \leq (1+(y_k,x_k)_B+\beta\alpha_k\|p_k\|^2/(2M\|B^{-1}\|))^2$$

from which now (3.6) immediately follows by taking square roots and using (3.9), (3.10).

Combining (3.5), (3.6) we obtain with (2.2b) the inequalities

$$(3.11) \qquad f(x_{k+1})-f(x_k) \;\geq\; (r_k,x_{k+1}-x_k) \;\geq\; M\|x_{k+1}-x_k\|^2 \;\geq\; 0.$$

In order to show (3.4) we estimate using (3.9), (3.10) and (2.2c)

$$(r_k, x_{k+1} - x_k) \;=\; [(r_k, x_k)(1 - \|x_k + \alpha_k p_k\|_B) + \alpha_k (r_k, p_k)] / \|x_k + \alpha_k p_k\|_B$$

(3.12) $$\geq \; \alpha_k \beta |p_k\|^2/(2L)$$

where L is an upper bound for $\|x_k + \alpha_k p_k\|_B$ on $U(x_0)$. The proof of the lemma is now complete.∎

In order to justify the choice of the Goldstein-Armijo stepsize rule instead of the constant α_k as in Lemma 3.1 we note that it may be shown as in [13] that

$$\|x_{k+1} - (x_k + \alpha_k p_k)\| \;=\; O(\alpha_k^2),$$

while (3.11), (3.12) yield an estimate linear in α_k. The proof of the first part of Theorem 2.10 is now an immediate consequence of (3.4), (3.11).

It remains to show the asymptotically quadratic convergence. In $U(x_0)$ the matrix D_k in (2.4b) is regular as a 'bordered' matrix. We next recall (cf. [6]) the expression for the derivative of an iteration function Φ as in (2.4).

The derivative of $\Phi(x) = y(x)/\|y(x)\|_B$, $y \in C^1$, is given by

(3.13) $$\Phi'(x) \;=\; y'(x)/\|y(x)\|_B P_y,$$

where $P_z = E_n - zz^T B/\|z\|_B^2$ is the orthogonal projector on $\{z\}^\perp$.

Now we show $\Phi'(x_0) = 0$ from which the quadratic convergence follows using Lemma 10.1.7 in [16].

Lemma ***3.M Under the assumptions of Theorem*** 2.10 ***the iteration function Φ of algorithm (2.4) satisfies***

$$\Phi'(x_0) \;=\; 0\,.$$

Proof. We note that in (2.4c) j=O will be chosen asymptotically and that then $\Phi(x)$ may be rewritten as (cf. (3.8b))

$$\Phi(x) \;=\; \rho y(x)/\|y(x)\|_B,\; y(x) \;=\; x - H(x)(\nabla f(x) - \lambda(x)Bx).$$

The regularity of D_0, (3.8a) and Lemma 10.2.1 in [16] yield

$$y'(x_0) = E_n - H_0(F_0 - \lambda_0 B) = -\, b_0 x_0^T B.$$

$y(x_0)=x_0$ and (3.13) then finally give

$$\Phi'(x_0) = -\rho b_0 x_0{}^T BP_{x_0} = 0.1$$

4. An Algorithm for Variational Inequalities

In this section we consider problem (1.2), (1.3). We present a globally convergent algorithm in the sense that it is not necessary as in Theorem 2.10 to choose x_1 in a sufficiently small neighborhood of a local maximum. Thus the following algorithm and theorem also generalize those of section 2.

We look for local maxima of the functional f defined on $H=\mathbb{R}^n$ over the set $K\cap\partial S_\rho$, ∂S_ρ as in (1.1), with g as in (2.2a) and K a discrete analog of (1.3):

$$K=\{x\in\mathbb{R}^n, x_i\geq 0,\ i\in J_1, x_i\leq 0,\ i\in J_2\},$$

$$J_1, J_2\subset\{1,\dots,n\}, J_1=\{i_1,\dots,i_{n_1}\}, J_2=\{j_1,\dots,j_{n_2}\}.$$

We introduce some further notation (cf. [13]). Let $G=(g_1,\dots,g_{n_1+n_2})$, where $g_k=e_{i_k}$, $k=1,\dots,n_1$, $g_{n_1+k}=e_{j_k}$, $k=1,\dots,n_2$, $e_i\in\mathbb{R}^n$ the i-th unit vector. Then K in (4.1) may be rewritten as

(4.2) $$K = \{x\in\mathbb{R}^n,\ G^Tx\geq 0\}.$$

For any $x\in\mathbb{R}^n$ let $I(x)=\{i\in\{1,\dots,2n\},\ g_i{}^Tx=0\}$ and define $G_I=(g_i)_{i\in I}$, $Q_I=E_n-G_IG_I{}^T$. For $x=x_k$ denote $I_k=I(x_k)$, $G_k=G_{I_k}$ and Q_k analogously. We can now define

The algorithm for variational inequalities

(4.3) Let $x_1\in K\cap\partial S_\rho$ be arbitrary. Set k=l and $\mu_k=0$, $\mu_k\in\{0,1\}$.

1. Determine I_k and $u_k=r_k-\lambda_kBx_k$, $\lambda_k=r_k{}^Tx_k/\rho^2$. Terminate the iteration if $G_k{}^Tu_k\leq 0$ and $\|Q_ku_k\|=0$.

2. Compute $|u_{kj}|=\max\{|u_{ki}|,\ (G_k{}^Tu_k)_i>0\}$. If $\{(Q_ku_k,r_k)<|u_{kj}|\ \|Q_ku_k\|$ and $\mu_k=0\}$ or $\|Q_ku_k\|=0$ then set $\tilde{I}_k=I_k-\{j\}$ and determine $\tilde{Q}_k$. Otherwise set $\tilde{I}_k=I_k$, $\tilde{Q}_k=Q_k$

3. Rcplacc $F_k-\lambda_k B$ in (2.4b) by $F_k-\lambda_k B-\tau_k E_n$, whcre $\tau_k=\max\{0,\delta+\sigma_k\}$ and σ_k is thc largcst cigcnvaluc of $F_k-\lambda_k B$ on $\{x_k\}^{\perp}\cap\{x\in\mathbb{R}^n, \tilde{Q}_k x=x\}$, $\delta>0$ a givcn constant. Compute p_k *as* thc dircction vcctor given by (2.4a) but in thc variables x_{ki} with $(\tilde{Q}_k)_{ii}=1$ (the frcc variables) only, fixing the others.

4. Dctcrmine the maximal admissible stcplcngth $\bar{\alpha}_k$ and the stcplcngth $\tilde{\alpha}_k$ as in (2.4) and sct

$$x_{k+1} = \rho(x_k+\alpha_k p_k)/\|x_k+\alpha_k p_k\|_B,$$

whcre $\alpha_k=\min\{\bar{\alpha}_k,\tilde{\alpha}_k\}$.

Thcorcm 4.4. ***Let the assumptions (2.2) be satisfied for problem (1.2). Assume that the set***

$$\Gamma = \{x^*\in K\cap\partial S_\rho,\ \ G^{*T}x^*\le 0,\ \ \|Q^*x^*\|=0\}$$

is finite and that $G^{*T}x^*<0$ for all $x^*\in\Gamma$ and $0<\delta<-\sigma^*$ (cf. 3. in (4.3)). Then the sequence $\{x_k\}$, $k=1,2,\ldots$, generated by algorithm (4.3) converges to a point $x^*\in\Gamma$. If $f\in C^3(U(x^*))$ then the asymptotic (Q-)order of convergence is two.

We first prove the analogue of Lemma 2.8 in the case $\rho=1$.

Lcmma 4.5. ***Let under the assumptions of Theorem 4.4 $x_k\in K\cap\partial S_\rho$ be generated 'by algorithm (4.3) with steplength $\alpha_k=\delta/(2M\mathrm{cond}(B))$ then $x_{k+1}\in K\cap\partial S_\rho$ and***

$$f(x_{k+1})-f(x_k) \ge \tilde{c}_k \begin{cases} \|p_k\|^2, & \text{if } \mu_k=1,\\ \max\{\|p_k\|, |u_{kj}|\}^2 & \textit{otherwise}, \end{cases} \tag{4.6}$$

where $\tilde{c}_k=c_1>0$ for $\mu_{k+1}=0$ and $\tilde{c}_k=c_2\bar{\alpha}_k$, $c_2>0$, for $\mu_{k+1}=1$.

Proof: The proof of (4.6) in the case $\mu_k=1$ follows closely the lines of the proof of Lemma 2.8. It is therefore not necessary here to give the -details. We remark only that p_k in (4.3) satisfies

$$(x_k,p_k)_B=0, \qquad \tilde{Q}_k(F_k-\lambda_k B-\tau_k E_n)\tilde{Q}_k p_k=-\tilde{Q}_k r_k \tag{4.7}$$

and the analogue of (3.9) holds with β replaced by 6. Let now $\mu_k=0$ and $\tilde{I}_k\ne I_k$. Since the analogue to (3.8b) shows that $\tilde{Q}_k r_k$ in (4.7) may be replaced by $\tilde{Q}_k u_k$, which contains the component u_{kj}, the assumptions of Theorem 4.4 assure that for a positive constant $c\|p_k\|\ge\|\tilde{Q}_k u_k\|\ge|u_{kj}|$. This provcs (4.6) for $\tilde{I}_k\ne I_k$. If $\tilde{I}_k=I_k$ then the strategy in 2. of (4.3) guarantccs that $\|Q_k r_k\|\ge|u_{kj}|$ whilc (4.7) givcs $c\|p_k\|\ge\|Q_k r_k\|$. This complctcs the proof of the lcmma.∎

The proof of thc first part of Thcorcm 4.4 nced also not be given in dctail hcrc sincc it follows from combining the argumcnts of thc proofs of Thcorcm 2.10 above and Thcorcm 3.1 in [13]. This shows that, for all sufficicntly largc k, $I_k = I(x^*)$, $x^* \in \Gamma$ and $\tau_k = 0$. Thus the stcplcngth α_k will finally be chosen equal to 1 and thc asymptotically quadratic convergence thcn follows as in the proof of Theorem 2.10.

Remark 4.8 For a practical application of algorithm (4.3) a way of choosing the regularization parameter τ_k has to be given. For a more gencral class of optimization problcms a proccdurc for this purpose is described in [20].

5. Path Following in Turning Point Problcms

In this section we consider the same class of problcms as in [6], namely the nonlinear eigenvalue problem (cf. (2.7a))

$$\text{(5.1)} \qquad \lambda L(u) = \varphi(x,u(x)), \quad x\in\Omega, \qquad u(x)=0, \quad x\in\partial\Omega,$$

where $\lambda\in\mathbb{R}$, DO, and L is a uniformly elliptic formally selfadjoint differential operator on the bounded domain $\Omega\subset\mathbb{R}^n$. Generalizations, for example, to higher order differential opcrators or other boundary conditions are possible. Conditions (2.2) have to be satisfied in the continuous case and for the discretization. We shall rcstrict ourselves to the example (cf. e.g. [4])

$$\text{(5.2)} \qquad L(u) = - \text{A u}, \qquad \varphi(x,u) = \exp(u/(1+ \text{EU})), \quad \varepsilon\geq 0.$$

and $N=2$. For $\varepsilon = 0$ (5.1), (5.2) is usually called Bratu's problem.

Therc has bcen a great interest in the numerical solution of similar problems, see, for example, thc papcrs mcntioncd in scction 5.6 of [14]. For thcorctical results on problcms of the type (5.1) sce, for cxample, [5, 7,191. It is well-known that (5.1), (5.2) has a solution diagram as shown in Fig. 1 in dimensions N= 1, 2. The points marked in the figure represent for $\varepsilon < \varepsilon^*$ one or two (quadratic) simple turning points rcspectively a nonsimple turning point for $\varepsilon = \varepsilon^*$.

Thc problcm of following thc solution branch and also thc problem of dctcrmining the simple rcspcctivcly the nonsimplc turning points numerically prcsents in principle no difficulties (cf. [2,14,21]). However, using e.g. Keller's pseudo-arclength-continuation technique the stepsize has to be suitably controlled near the limit point and the question of efficiency arises in particular if the linear systems are solved by elimination methods. In [2] a multigrid(MG)-method was suggested for the approximate solution of (5.1), (5.2). The pseudo-arclength normalization was added (cf. [8]) and the resulting system was solved by block-elimination as utilized also in Remark 2.6. Hence a differential operator was discretized, which becomes singular in the turning point. The corresponding singularity of the discrete operator on one of the grids used in the MG-method made it necessary to modify this algorithm considerably in order to be able to pass the limit point.

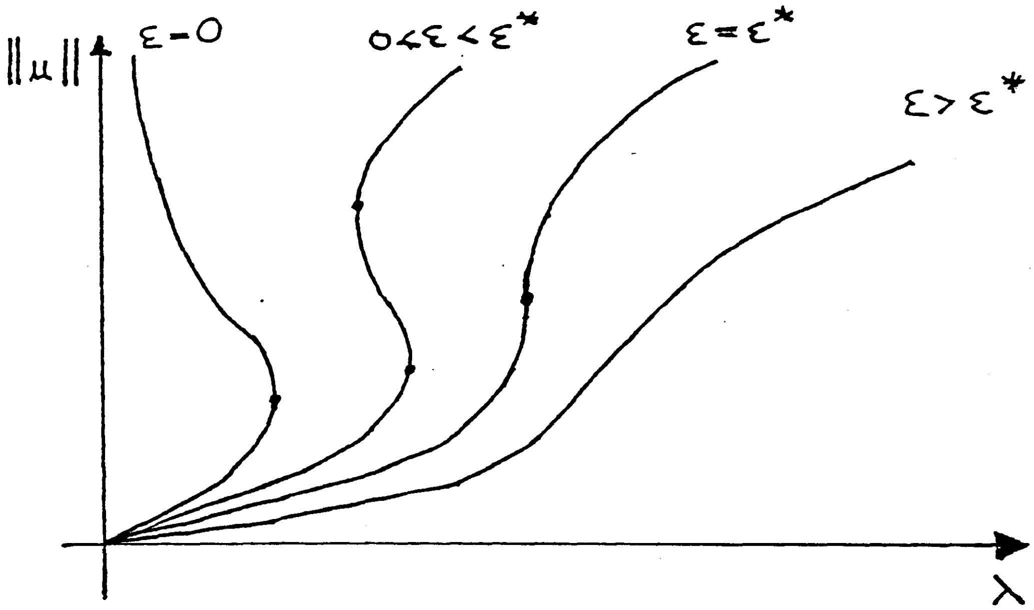

Fig. 1. Solution diagram for problem (5.1). (5.2) for different values of ε.

These modifications may not have been necessary, if instead the inflated system would have been treated directly. The resulting system has a regular matrix in the neighborhood of solutions. However, the matrix is not definite on the whole space, so that it is open how the MG-method would perform. This question will be investigated in the future. For an application of MG using Rayleigh-quotient iteration to the linear eigenvalue problem cf. [11].

It is a well-known procedure to use a norm of u as a continuation parameter and a numerical method for this is, for example, the Picard iteration of [6]. The algorithms of sections 2 and 4 can be used analogously. They have the advantage of quadratic convergence while Fast-Poisson-Solvers in the special case $L= -\Delta$ could in general not be utilized. It should, however, as pointed out above, be possible to use MG-algorithms.

We compare now algorithm (2.4) and that of [6] on the above problem. Since it is not our aim to compute the solution to a high accuracy we have chosen a low order finite element method on a relatively coarse mesh. Problem (5.1), (5.2) may be written in the variational form

$$\lambda(\nabla g(u),v)=(\nabla f(u),v),\quad \forall v\in H_0^1(\Omega),$$

(5.3)

$$g(u)=\tfrac{1}{2}\int_\Omega (u_x^2+u_y^2)\,dxdy,\quad f(u)=\int_\Omega \exp(u/(1+\varepsilon u))\,dxdy.$$

Ω was taken as the unit square and linear finite elements were used on the standard triangulation obtained from a square mesh with meshwidth h. f was evaluated by numerical integration with weights $h^2/6$ and the midpoints of the edges of a triangle as integration points. This gave rise to the usual five-point difference matrix B and a seven-band matrix A. Table 1 shows the results for two values of $\varepsilon<\varepsilon^*$.

ε	ρ	λ	Alg.(2.4)	Picard
0.0	30	6.712380	4 (4)	9 (17)
0.0	36	6.910483	2 (3)	8 (16)
0.0	42	6.882701	2 (3)	8 (16)
0.0	48	6.681038	2 (3)	9 (16)
0.2	72	9.278187	3 (3)	10 (18)
0.2	80	9.291875	2 (3)	9 (18)
0.2	88	9.265477	2 (3)	9 (18)
0.2	96	9.211836	2 (3)	8 (18)
0.2	360	7.341984	2 (3)	9 (20)
0.2	440	7.237885	2 (3)	9 (20)
0.2	520	7.230358	2 (3)	9 (20)
0.2	600	7.285922	2 (3)	8 (20)

Table 1. Computed points on the solution branch for problem (5.1). (5.2) near the turning points and necessary number of iterations for different algorithms.

For either method the number of iterations is given required to compute the solution to about eight decimal places with the number of iterations for maximal accuracy (Double precision FORTRAN on an IBM 370-168) given in parantheses. The starting vector for $\rho = 30$ and for both algorithms was $x_0 = e/\|e\|_B$, $e = (1,\ldots,1)^T \in \mathbb{R}^n$, $n = ((1-h)/h)^2$, $h = 1/12$. The approximate solution for each ρ-value was then, after normalization, used as starting guess for the next $\rho(\varepsilon)$-value. Algorithm (2.4) could in each case be used with $\alpha_k = 1$. The linear system for the symmetric but in general indefinite matrix D_k in (2.4b) may bc solved, for example, by any conjugate gradient method applicable to such problems (see, for example, [3]) and even special elimination procedures arc easy to derive. We used algorithm SYMMLQ ([17]) which without any scaling or preconditioning needed about 35 iterations to solve the system in each step.

The iterates of our algorithm converged quadratically from the beginning. The steps in ρ for (2.4) could be chosen large as the results show, but not arbitrarily large, while the Picard iteration did not seem to have similar restrictions. So an alternative to damping in (2.4) could be to first execute some Picard steps and then to use algorithm (2.4) with stepsize 1.

In this section we have seen that algorithm (2.4) may be used very efficiently in the following of solution branches for problems of the type (5.1). For more general bifurcation problems a natural procedure would be to use alternately continuation with respect to λ or to the norm of x (cf. section 6) switching when the steplength in one of the methods has to be chosen below a suitable tolerance. The use of MG-methods may be possible, however, conjugate gradient algorithms provide an efficient and generally applicable procedure for the solution of the linear systems.

6. Path Following in Bifurcation Problems for Variational Inequalities

In this section we again restrict the numerical computations to a simple but illustrative example. We apply algorithm (4.3) to the discretization used in [13] of the buckling problem for an axially compressed beam with lateral supports. The variational inequality is

$$\lambda(\nabla g(u_0), u-u_0) \geq (\nabla f(u_0), u-u_0), \qquad \forall u \in K,$$

$$f(u) = \int_0^1 [(1+u'^2)^{1/2} - 1]\,dx, \qquad g(u) = 1/2 \int_0^1 u''^2\,dx, \tag{6.1}$$

$$K = \{u \in H_0^2[0,1],\ u(C) \geq 0,\ u(D) \leq 0\}.$$

Hermite cubic finite elements on an equidistant grid of width h and suitable numerical integration are used yielding the discrete functionals f_h, g_h (cf. [13]). Of physical interest are the solutions u_h branching from the trivial solution at the largest eigenvalue λ_{h1} with eigenvector u_{h1} of the linearized problem.

To our knowledge no reasonably efficient algorithms are available which are globally convergent to u_{h1} if $K \neq H$, except in special cases (see, for example, Corollary 4.2 in [13]). In [10] a constructive existence proof for the restricted solutions has been given in which they are obtained as bifurcating solutions of a penalized version of the unrestricted problem $(K=H)$. In that paper, however, only eigenfunctions can be determined corresponding to eigenvalues which are smaller than the largest eigenvalue of the unrestricted problem for which the corresponding eigenfunction with suitably chosen sign is in the interior of K. Hence the physically interesting case is excluded.

We assume now that (λ_{h1}, u_{h1}) and the corresponding set of active constraints are known and try to follow the branch bifurcating from $(\lambda_{h1}, 0)$. In [13] it was suggested that augmented Lagrangian methods could advantageously be used for this purpose. The following results, however, show that algorithm (4.3) which here essentially reduces to (2.4) in the subspace of the free variables is the most efficient method among several algorithms. We compared it with SALMNA, an augmented Lagrangian type algorithm using Newton's method from the NPL-library and also part of the NAG-library. Another natural candidate for a comparison is X-continuation (see, for example, [S]) which in this case should not be inferior to pseudo-arclength-continuation:

Let (u^0, λ^0) on the branch be given. Compute $u_\lambda(u^0, \lambda^0)$ from

$$(F(u^0) - \lambda^0 B)u_\lambda = Bu^0.$$

Then set $u_0 = u^0 + (\lambda - \lambda^0)\dot{u}_\lambda$ and for $k = 0,1,....$ iterate according to

$$(F(u_k) - \lambda B)(u_{k+1} - u_k) = -\nabla f(u_k) + \lambda Bu_k.$$

Hence after an Euler predictor step several Newton steps are executed to compute the solution for the given λ. Finally, Picard iteration is applied here, too.

We have restricted the computations to the problem (6.1) with $C = \{1/3\}$, $D = \{2/3\}$.

Largest eigenvalue and corresponding eigenfunctions for this case, there are two symmetric eigenfunctions, have been computed analytically in [13]. Table 2 shows some typical results for $h = 1/24$. Again the number of iterations is given required to compute the solution to eight decimal places respectively to the maximal attainable accuracy. For SALMNA the numbers represent for the latter case only the number of second order derivative (function value and first order derivative) evaluations.

ρ	$\lambda_{h\rho}$	Alg. (4.3)	Picard	λ-cont.	SALMNA
1	.144644 E - 1	2 (3)	14 (28)	5 (6)	2 (9)
2	.139243 E - 1	2 (3)	15 (28)	4 (5)	2 (8)
10	.799582 E - 2	3 (4)	16 (30)	5 (6)	9 (14)
100	.103968 E - 2	3 (4)	14 (28)	7 (8)	70 (104)

Table 2. Computed points on the bifurcating branch for problem (6.1) and iteration counts for different algorithms.

For each algorithm the normalized eigenfunction of the linear eigenvalue problem was used as starting solution for $\rho = 1$ and the corresponding Rayleigh-quotient was used as starting value for the Lagrange multiplier in SALMNA. Then the solutions on the branch for the given sequence of p-values were computed by continuing analogously to $\rho = 2$, 10, 100. The corresponding X-values were used as the sequence for the X-continuation.

The results show that our method is also very efficient for following bifurcating branches of variational inequalities. The behavior of the Picard iteration is similar to that in section 5, while for X-continuation the convergence of the Newton iterates was not quadratic from the start which resulted in considerably more iterations especially for larger p-steps. The iteration counts for this method in Table 2 do not include the predictor step. Finally the performance of the general purpose routine SALMNA suggests that augmented Lagrangian methods are not able to compete with algorithm (4.3) for the special class of optimization problems considered here. By modifying the subroutine suitably it should, however, be possible to reduce the extremely high expense needed for larger p-steps.

For the solution of the linear systems again SYMMLQ was used which even after a scaling of the system needed more than n iterations. The number of iterations, however, was only slightly larger than that for the solution of the system in the Picard iteration which has a definite matrix. So this difficulty is caused by the unfavourable eigenvalue distribution for this fourth-order problem and, if conjugate gradient methods are to be used for the linear systems, a suitable preconditioning should be chosen to further reduce the necessary work.

Acknowledgement. The author sincerely thanks Professor G.H. Golub for the opportunity to spend six months in the Numerical Analysis group of Stanford University and for his support and encouragement during the writing of this paper. He also thanks the members of the NA-group for their help. particularly with the use of the various computing facilities.

References

1. Berger, M.S.: A bifurcation theory for nonlinear elliptic partial diffcrcntia! equations and related systems. In: Bifurcation theory and nonlinear eigenvalue problems. J.B. Keller and S. Antman, cds.. W.A. Benjamin Inc., New York 1969

2. Chan, T.F.C. and Keller, H.B.: Arclength continuation and multi-grid techniques for nonlinear elliptic eigenvalue problems. Techn. Rep. no. 197, Dept. of Comp. Science, Yale University 1981

3. Chandra, R.: Conjugate gradient methods for partial differential equations. Techn. Rep. no. 129. Dcpt of Comp. Science, Yale University 1978

4. Fradkin. L.J. and Wake, G.C.: The critical explosion parameter in the theory of thermal ignition. J. Inst. Math. App!. 20, 471-484 (1977)

5. Gelfand, I.M.: Some problems in the theory of quasilinear equations. AMS Translations 29, 295-381 (1963)

6. Gcorg, K.: On the convergence of an inverse iteration method for nonlinear elliptic eigenvalue problems. Numer. Math. 32, 69-74 (1979)

7. Keller, H.B.: Some positone problems suggested by nonlinear heat generation. In: Bifurcation theory and nonlinear eigenvalue problems. J.B. Keller and S. Antman, eds., W.A. Benjamin Inc., New York 1969

8: Keller. H.B.: Numerical solution of bifurcation and nonlinear cigcnvalue problems. In: Applications of bifurcation theory. P.H. Rabinowitz, ed., Academic Press, New York 1977

9. Kratochvil, A., Necas, J.: Gradient methods for the construction of Ljusternik- Schnirelmann critical values. Revue Francaise Automat Informat. Recherche Operationelle R 14, 43-54 (1980)

10. Kucera, M.: A new method for obtaining eigenvalues of variational inequalities based on bifurcation theory. Casopis pro pestovani matematiky **104**, 389-411 (1979)

11. McCormick, S.F.: A mesh refinement method for $Ax = \lambda Bx$. Math. Comp. 36, 485-498 (1981)

12. Miersemann, E.: Verzweigungsprobleme fur Variationsungleichungen. Math. Nachr. 65, 187-209 (1975)

13. Mittelmann, H.D.: Bifurcation problems for discrete variational inequalities. Math. Meth. App!. Sci. (to appear)

14. Mittelmann, H.D. and Weber. H.: Numerical methods for bifurcation problems - a survey and classification. In: Bifurcation problems and their numerical solution. H.D. Mittelmann and H. Weber. eds., ISNM 54, Birkhäuser-Verlag, Base! 1980

15. Necas, J.: Approximation methods for finding critica.! points of even functionals. Trudy Matem. Inst. A.N. SSSR **134**, 235-239 (1975)

16. Ortega, J.M. and Rheinboldt, W.C.: Iterative solution of nonlinear equations in several variables. Academic Press, New York and London 1970 .

17. Paige, C.C. and Saunders, MA.: Solution of sparse indefinite systems of linear equations SIAM J. Numer. Anal. 12. 617-629 (1975)

18. Peters, G. and Wilkinson, J. H.: Inverse iteration. ill-conditioned equations and Newton's method. SIAM Review 21. 339-360 (1979)

19. Rabinowitz, P.H.: Variational methods for nonlinear elliptic eigenvalue problems Indiana Univ. Math. Journal **23**, 729-754 (1974)

20. Spellucci, P.: Some convergence results for generalized gradient projection methods. Methods of Operations Research 36, Verlag Anton Hain, Königsstein 1980, pp. 271-280.

21. Spence, A. and Werner, B.: Non simple turning points and cusps. To appear as Techn. Rep., Inst. fur Angew. Mathem., Universität Hamburg 1981

22. Wilkinson, J.H.: The algebraic eigenvalue problem. Clarendon Press, Oxford 1965