MONOTONE CONVERGENT METHODS FOR A VARIATIONAL INEQUALITY

DAVID E. STEWART[1] AND STEPHEN J. WRIGHT[2]

ABSTRACT. Penalty and path-following methods have been used for solving finite-dimensional quadratic programmes. The intention here is to apply such techniques to an infinite-dimensional problem, namely a one-sided obstacle problem, and to develop a method for solving the problem in an infinite-dimensional setting. The numerical methods developed in the infinite-dimensional context, so the convergence rate of the discretisations are (in some sense) independent to the size of the finite-dimensional approximation. These methods are shown to be convergent in appropriate Banach spaces by means of a monotonicity result for the iterates of the associated Newton method. This montonicity carries over to finite-dimensional discretisations for a large class of methods. The overall numerical method developed is based on an exterior penalty function, and some numerical results have been obtained.

1. Introduction.

Variational inequalities are problems that often arise in connection with free-boundary problems, contact problems and elastic-plastic problems[2]. A well-known example of such a problem is the so-called "obstacle problem" [**2, pp. 3, 104ff**]. The idea is to model an elastic sheet which is suspended over some terrain. The sheet is assumed to be supported at its edges, but it might also be supported by the ground, though what over region the sheet is in contact with the ground is unknown. Where the sheet is suspended, the usual equations apply, though over the region of contact, a frictionless force is exterted on the sheet by the ground to keep it from sinking into the ground. A linearised version of this can be represented by the equations below, where $u(x,y)$ is the downward vertical displacement of the sheet at co-ordinates (x,y).

$$-\Delta u = f(x,y) \qquad \text{where } u(x,y) < b(x,y)$$
$$-\Delta u \le f(x,y) \qquad \text{where } u(x,y) = b(x,y)$$

where $b(x,y)$ is the downwards displacement of the ground and $f(x,y)$ is the downward force applied per unit area of the sheet. Typically this force would be gravity.

Such problems can be considered to be infinite-dimensional versions of quadratic programmes as they can be reformulated as minimisation problems in Hilbert spaces. For example, the above obstacle problem can be re-formulated as

$$\min_u \int_\Omega \left[\frac{1}{2}|\nabla u|^2 - fu\right] dV$$

where the minimisation is done over all u where $u(x,y) \ge b(x,y)$ for all (x,y).

Numerically such methods can be discretised and solved as large quadratic programmes. However, in spite of the recent advances in such methods, the guaranteed convergence rate of such methods are highly dependent on the dimension of

[1] School of Mathematical Sciences, Australian National University, GPO Box 4, Canberra, ACT

the problem. Furthermore, the rates of convergence go to zero as this dimension is increased.

Here an alternative approach is used to find "grid-independent" methods by developing algorithms in the appropriate Hilbert space, which can then be transformed into numerical methods. The convergence rates of these methods will approach that of the infinite-dimensional algorithm as the grid is refined and the approximations made more accurate[1]. Further the convergence is monotone in the sense that successive iterates $u^{(k)} \in H_0^1$ are decreasing in k *as functions*. That is, $u^{(k+1)}(x) \leq u^{(k)}(x)$ for all $x \in \Omega$.

The resulting algorithms are not the fastest available methods, as truncated SOR type methods (such as [2]) seem to be able to solve the problem in about the time it takes to solve a single linear elliptic PDE. However, the structure that is evident with the approach of this paper is not utilised or preserved by such methods.

2. Main problem and formulation.

The basic problem to be solved here is

$$\min_u \int_\Omega \left[\frac{1}{2}|\nabla u|^2 - c(x)u\right] dV(x) \tag{2.1}$$

over $u \in H_0^1(\Omega)$ and $u(x) \leq b(x)$ for all x in Ω. We assume that b is Lipschitz continuous and $b \geq 0$ on $\partial\Omega$, and that c belongs to $L^2(\Omega)$. The region Ω is assumed to be an open subset of $\mathbf{R}^n$ with a Lipschitz boundary, and that $n \leq 4$. This minimisation problem is equivalent to the variational inequality

$$\int_\Omega [\nabla u \cdot \nabla(v-u) - c(v-u)]\, dV \geq 0$$
$$u(x) \geq b(x) \qquad \text{for all } x \in \Omega$$

for all $v \in H_0^1$, $v(x) \geq b(x)$ for all $x \in \Omega$. The dependence on $x \in \Omega$ is usually suppressed in what follows, unless it needs to be explicitly referred to. Where solutions to PDEs are referred to below, it is assumed that they are all in $H_0^1(\Omega)$. That is, the homogenous boundary conditions $u|\partial\Omega = 0$ hold.

Denote the functional in (2.1) by $J(u)$. That solution exists for this problem follows from the fact that the functional J is a convex and continuous function $H_0^1(\Omega) \to \mathbf{R}$, and hence is weakly lower semi-continuous (LSC). It is coercive on H_0^1, using the norm

$$\|u\|_{H_0^1} = \sqrt{\int_\Omega |\nabla u|^2\, dV}.$$

(Coercivity is easy to see from $J(u) \geq \|u\|_{H_0^1}^2 - \|c\|_{H^{-1}}\|u\|_{H_0^1}$.) The set $\{u \in H_0^1(\Omega) \mid u \leq b\}$ is nonempty. If $\bar{u}$ is a given element of this set, then $\{u \in H_0^1(\Omega) \mid u \leq b,\ J(u) \leq J(\bar{u})\}$ is nonempty, closed and bounded; it is therefore weakly compact, and since the functional J is weakly LSC, the minimum of J subject to $u \leq b$ must exist. For the solution to satisfy the first order "necessary" conditions, it is also necessary that J is *proper*. That is, $J(u) < +\infty$ for some $u \in H_0^1$, $u \leq b$. If b is Lipschitz, then $u(x) = \min(0, b(x))$ suffices.

Since (2.1) is a difficult problem to solve directly, we consider a penalty functional

$$J_\mu(u) = \int_\Omega \left[\frac{1}{2}|\nabla u|^2 - cu + \mu\Phi(x,u(x))\right] dV. \tag{2.2}$$

In particular, there are two penalty functionals that we expect to use: an *interior penalty* or *barrier* $\Phi(x,u) = -\ln(b(x) - u)$, and an *exterior penalty* $\Phi(x,u) = (u - b(x))_+^2$. In the former case, the original problem is recovered as $\mu \downarrow 0$, and in the latter, as $\mu \to +\infty$. In both cases, a number of crucially important properties hold. Let $(\cdot)_u$ denote differentiation with respect to u. Then we require of Φ that

$$\begin{array}{ll} \Phi(\cdot,u) & \text{is bounded below by an integrable function, for all feasible } u \\ & \Phi_u,\ \Phi_{uu} \geq 0. \\ & \Phi_{uu} \text{ is non-decreasing.} \end{array} \tag{2.3}$$

The necessary conditions for a minimum of J_μ are that

$$-\Delta u - c + \mu\Phi_u(x,u) = 0 \qquad \text{in } \Omega \tag{2.4}$$

and $u = 0$ on $\partial\Omega$. Solutions to this PDE exist for sensible Φ by the same sort of arguments as were used to show that sensible minima of J exist. The functional J_μ is coercive, convex and LSC, and so it is weakly LSC on a weakly compact set. Provided J_μ is proper (that is $J_\mu(u) < +\infty$ for some $u \in H_0^1$), a finite minimum exists and this minimum satisfies the above necessary conditions (2.4). To solve this problem, a Newton method is used, which has some important properties which we discuss in the next section.

3. The Newton method.

The iterates of the Newton method for (2.4) have a remarkable monotonicity property, which can be used to prove a global convergence property. This monotonicity property is the basis of the results in later sections.

Let $\phi(x,u) = \Phi_u(x,u)$. Given a starting function u, the Newton method for (2.4) computes $u + \delta u$ where

$$-\Delta(u+\delta u) - c + \mu\phi(x,u) + \mu\phi_u(x,u)\delta u = 0. \tag{3.1}$$

Thus

$$-\Delta(\delta u) + \mu\phi_u(x,u)\delta u = -\left[-\Delta u - c + \mu\phi(x,u)\right] = f(x). \tag{3.2}$$

It should be noted that this linear PDE is solvable, even though $\mu\phi_u(x,u(x))$ may not belong to L^∞. The arguments to show this are along the lines whereby the orignal penalised problems where shown to have solutions. Consider the quadratic functional in δu:

$$\int_\Omega \left[\frac{1}{2}|\nabla \delta u|^2 + \mu\phi_u(x,u(x))(\delta u)^2 - f(x)\delta u\right] dV. \tag{3.3}$$

This functional is coercive, strictly convex (since $\phi_u \geq 0$), LSC and *proper* (that is, the functional is finite for some $\delta u \in H_0^1$; for example, take $\delta u \equiv 0$). Hence the functional (3.3) has a unique minimiser in H_0^1. This minimiser is the solution to the PDE (3.2).

If $-\Delta u - c + \mu\phi(x,u) \geq 0$ and $\phi_u \geq 0$, then by the maximum principle for elliptic PDEs[7], $\delta u \leq 0$. Furthermore,

$$\begin{aligned}-\Delta(u+\delta u) + \mu\phi(x,u+\delta u) - c &= \mu\phi(x,u+\delta u) - \mu\phi(x,u) - \mu\phi_u(x,u)\delta u \\ &\geq 0\end{aligned}$$

since $\phi_u = \Phi_{uu}$ is non-decreasing.

This gives the desired monotonicity property for the iterates of Newton's method: If we denote the successive iterates of Newton's method by $u^{(k)}$, $k = 0, 1, \ldots$, then if $-\Delta u^{(0)} - c + \mu\phi(x, u^{(0)}) \geq 0$, then

$$u^{(0)} \geq u^{(1)} \geq \cdots \geq u^{(k)} \geq u^{(k+1)} \geq \ldots$$

Now we show that provided $\phi(x, u(x))$ and $\phi_u(x, u(x))$ are in $L^2(\Omega)$, then the iterates converge in L^2. The trickiest part of the proof is just to show that the Newton iterates $u^{(k)}$ are bounded in L^2. To obtain such a bound, let w solve the problem

$$-\Delta w = -|c| - \mu\phi(x,0). \tag{3.4}$$

Then for any $u^{(k)}$, as $\phi_u(x,u) \geq 0$ for all $x \in \Omega$ and u, it follows that

$$\begin{aligned}-\Delta w + \mu\phi_u(x,u^{(k)})w &\leq -\Delta w \\ &= -|c| - \mu\phi(x,0) \\ &\leq c - \mu\phi(x,0).\end{aligned}$$

Note that $\phi(x,0) \in L^2$ since $0 \leq \phi(x,0) \leq \phi(x,u^{(k)}(x))$ and $\phi(x,u^{(k)}(x))$ is in L^2. Thus w is (in fact) in H^2.

Now

$$\begin{aligned}-\Delta u^{(k+1)} + \mu\phi_u(x,u^{(k)})u^{(k+1)} &= c - \mu\phi(x,u^{(k)}) + \mu\phi_u(x,u^{(k)})u^{(k)} \\ &\geq c - \mu\phi(x,0)\end{aligned}$$

since $\phi(x,0) \geq \phi(x,u^{(k)}) - \phi_u(x,u^{(k)})u^{(k)}$. Hence by the maximum principle, $u^{(k+1)} \geq w$ for all k.

Thus

$$u^{(0)} \geq u^{(1)} \geq \cdots \geq u^{(k)} \geq u^{(k+1)} \geq \cdots \geq w.$$

As $w \in H_0^1$, it is also in L^2. Hence by the monotone convergence theorem for Lebesgue integration, $u^{(k)}$ has a (strong) limit $\bar{u}$ in L^2. This can be improved by using the Sobolev imbedding theorem[**2**,**pp. 80, 92**].

Since the lower bound w is in H^2, it must be in L^p for $p \leq 2n/(n-4)$ if $n > 4$, and any $p < \infty$ for $n \leq 4$[**2, p. 92**]. (If $n < 4$ then w must be continuous. For

$n \leq 8$, w is in L^4. This will be used later.) Thus if $u^{(0)} \in L^p$ with the above restrictions on p, the monotonicity shows that $u^{(k)} \to \bar{u}$ in L^p.

The question to answer next is whether $\bar{u}$ is a solution of the original problem. Now

$$-\Delta u^{(k+1)} = c - \mu\phi(x, u^{(k)}) - \mu\phi_u(x, u^{(k)})(u^{(k+1)} - u^{(k)}).$$

Provided $u^{(0)}$, $\phi(x, u^{(0)})$ and $\phi_u(x, u^{(0)})$ are all in L^4, it follows that the right-hand side converges in L^2 to $c - \mu\phi(x, \bar{u})$, and hence $u^{(k+1)}$ converge strongly in H_0^1, also to $\bar{u}$. Hence, $\bar{u}$ is indeed a solution to the problem (2.2).

Rate of convergence.

Now that strong convergence of $u^{(k)} \to \bar{u}$ in H_0^1 has been proven, we can investigate the speed with which this convergence occurs. As we are using a Newton method, we would expect quadratic onvergence if ϕ is a C^2 function and there is no "degeneracy" in the problem.

Quadratic convergence for smooth ϕ can indeed be shown: using

$$-\Delta u^{(k+1)} - c + \mu\phi(x, u^{(k)}) + \mu\phi_u(x, u^{(k)})(u^{(k+1)} - u^{(k)}) = 0$$
$$-\Delta\bar{u} - c + \mu\phi(x, \bar{u}) = 0$$

we get

$$-\Delta(u^{(k+1)} - \bar{u}) + \mu\left[\phi(x, u^{(k)}) + \phi_u(x, u^{(k)})(u^{(k+1)} - u^{(k)}) - \phi(x, \bar{u})\right] = 0.$$

Now

$$\phi(x, u^{(k)}) + \phi_u(x, u^{(k)})(u^{(k+1)} - u^{(k)}) - \phi(x, \bar{u}) =$$
$$\phi_u(x, u^{(k)})(u^{(k+1)} - \bar{u}) - \left[\phi(x, \bar{u}) - \phi(x, u^{(k)}) - \phi_u(x, u^{(k)})(\bar{u} - u^{(k)})\right].$$

Hence

$$-\Delta(u^{(k+1)} - \bar{u}) + \mu\phi_u(x, u^{(k)})(u^{(k+1)} - \bar{u}) =$$
$$\mu\left[\phi(x, \bar{u}) - \left(\phi(x, u^{(k)}) + \phi_u(x, u^{(k)})(\bar{u} - u^{(k)})\right)\right].$$

Let $M_k = \frac{1}{2}\sup_{x\in\Omega,\, \bar{u}(x)\leq v\leq u^{(k)}(x)} \phi_{uu}(x, v)$. Note that M_k is non-increasing in k since $u^{(k+1)} \leq u^{(k)}$. Thus, provided ϕ has bounded second derivatives, the right-hand side of the above equation for $u^{(k+1)} - \bar{u}$ can be bounded pointwise by

$$\mu M_k(\bar{u} - u^{(k)})^2.$$

This can be used to obtain bounds on $\|u^{(k+1)} - \bar{u}\|_{H_0^1}$ by multiplying by $u^{(k+1)} - \bar{u}$ and integrating over Ω. Using integration by parts,

$$\int_\Omega \left[|\nabla(u^{(k+1)} - \bar{u})|^2 + \mu\phi_u(x, u^{(k)})(u^{(k+1)} - \bar{u})^2\right] dV \leq$$
$$\mu\int_\Omega M_k(u^{(k+1)} - \bar{u})(\bar{u} - u^{(k)})^2\, dV \leq$$
$$\mu M_k\|u^{(k+1)} - \bar{u}\|_{L^2}\|(\bar{u} - u^{(k)})^2\|_{L^2}.$$

The left-hand side of this inequality is $\geq \int_\Omega |\nabla(u^{(k+1)} - \bar{u})|^2\, dV = \|u^{(k+1)} - \bar{u}\|^2_{H_0^1}$. If $k_{\Omega,2}$ denotes the imbedding constant of $H_0^1(\Omega)$ in $L^2(\Omega)$, then

$$\|u^{(k+1)} - \bar{u}\|_{L^2} \leq k_{\Omega,2}\|u^{(k+1)} - \bar{u}\|_{H_0^1}.$$

Substituting this into the inequality for the above integral gives

$$\|u^{(k+1)} - \bar{u}\|_{H_0^1} \leq \mu M_k k_{\Omega,2}\|(\bar{u} - u^{(k)})^2\|_{L^2}$$

after dividing by $\|u^{(k+1)} - \bar{u}\|_{H_0^1}$.

If $n \leq 4$, then $H_0^1(\Omega) \subset L^4(\Omega)$. Let $k_{\Omega,4}$ be the imbedding constant; that is, $\|u\|_{L^4} \leq k_{\Omega,4}\|u\|_{H_0^1}$ for all $u \in H_0^1(\Omega)$. Then,

$$\|u^{(k+1)} - \bar{u}\|_{H_0^1} \leq \mu M_k k_{\Omega,2}\|u^{(k)} - \bar{u}\|^2_{L^4} \leq$$
$$\leq \mu M_k k_{\Omega,2} k^2_{\Omega,4}\|u^{(k)} - \bar{u}\|^2_{H_0^1}.$$

Thus convergence is quadratic for fixed μ and smooth ϕ.

4. Following the "central path".

The "central path" is the set of (μ, u_μ) in $\mathbf{R}_+ \times H_0^1(\Omega)$ which minimises (2.2). The first result of this section is that if $0 < \mu_1 < \mu_2$ then $u_{\mu_1} \geq u_{\mu_2}$. This follows easily from the results of the previous section by noting that

$$-\Delta u_{\mu_1} - c + \mu_2\phi(x, u_{\mu_1}) \geq -\Delta u_{\mu_1} - c + \mu_1\phi(x, u_{\mu_1}) = 0$$

and so by the monotonicity of the Newton iterates, $u_{\mu_1} \geq u_{\mu_2}$.

Log barrier functional.

For the log barrier functional, taking $\mu \downarrow 0$ gives a monotonic increasing sequence of elements in H_0^1. If we change $\Phi(x,u)$ to be the function $-\ln((b(x) - u)/(b(x) - w(x))$ then the functional J_μ is monotonic decreasing in μ for any u with $w \leq u < b$. This change to the functional makes no difference to the algorithm (since $\phi(x,u) = \Phi_u(x,u) = 1/(b(x) - u)$ regardless of the factor $1/(b(x) - w(x))$). Thus, if $J_{\mu_0}(u_{\mu_0}) < \infty$, then $J(u_\mu) \leq J_\mu(u_\mu) \leq J_\mu(u_{\mu_0}) \leq J_{\mu_0}(u_{\mu_0}) < \infty$ for $\mu_0 \geq \mu > 0$. Thus as $\mu \downarrow 0$, the functions u_μ belong to a bounded subset of $H_0^1(\Omega)$, and so there is a weakly convergent subsequence. Now $b \geq u_\mu \geq u_{\mu_0}$ so by the arguments of the previous section, u_μ converges strongly in L^2 to $\bar{u}$. Thus u_μ can only have one weak limit in H_0^1, which is $\bar{u}$.

Now J is a weakly lower semi-continuous functional on H_0^1, so

$$J(\bar{u}) \leq \lim_{\mu\downarrow 0} J(u_\mu) < \infty.$$

The problem now is to show that $\bar{u}$ minimises J over all functions u in H_0^1 where $u \leq b$. As b isLipschitz continuous, given any such u and any $\epsilon > 0$, there is a $\hat{u} \in H_0^1$ which satisfies $\Phi(x, \hat{u}) \in L^1$ and $J(\hat{u}) \leq J(u) + \epsilon$. Then as

$$J(u_\mu) \leq J_\mu(u_\mu) \leq J_\mu(\hat{u}) \leq J(\hat{u}) + \mu\|\Phi(\cdot, \hat{u}(\cdot))\|_{L^1} \leq$$
$$\leq J(u) + \epsilon + \mu\|\Phi(\cdot, \hat{u}(\cdot))\|_{L^1},$$

taking $\mu \downarrow 0$ gives

$$J(\bar{u}) \leq \lim_{\mu \downarrow 0} J(u_\mu) \leq J(u) + \epsilon.$$

Thus $\bar{u}$ minimises J over the set of feasible functions, and thus solves the variational inequality.

Following the central path using Newton corrections is, however, not easy to justify theoretically. The problem is that to ensure monotone convergence requires that the initial "guess" $\hat{u}^0_\mu$ satisfies $\hat{u}^0_\mu \geq u_\mu$. If the function u_{μ_0} is known and $\mu_0 > \mu$ then setting the initial "guess" to be $u^0_\mu = u_{\mu_0}$ will be an underestimate of u_μ. This means that the first Newton iterate u^1_μ will be an overestimate, and it is not clear by how much it will be an overestimate.

Taking "sufficiently small steps" would result in a convergent algorithm for finite-dimensional problems, however, if $n \geq 2$, bounded functions in the H^1_0 norm are not necessarily bounded functions in the supremum norm, and the Newton correction may be pointwise unbounded. In this case, *any* step of positive size in the direction of the indicated correction would result in the penalty function becoming pointwise unbounded, and even unbounded in L^1. It may be that in certain cases with $n \geq 2$ that the logarithmic barrier method loses the grid-independence property. The computed corrections may, in fact, belong to H^2, and because of this additional regularity the method may be provably "grid-independent" for $n \leq 2$, though the same problem would arise with $n = 3$ with regards to lack of pointwise bounds. Numerical evidence for $n = 2$ seems to indicate that the logarithmic barrier method gives good performance[8], which seems to indicate that computed corrections may well be bounded in H^2.

Exterior penalty function.

For the exterior penalty function, we take $\mu \to +\infty$. Now u_μ decreases with increasing μ. Let u^* be the solution of the variational inequality. Since $J(u_\mu) \leq J_\mu(u_\mu) \leq J_\mu(u^*) = J(u^*)$, it follows that $J(u_\mu)$ is uniformly bounded, and as J is coercive, u_μ must all lie in a bounded set in H^1_0. Thus there is a subsequence of u_μ that is weakly convergent in H^1_0, and since H^1_0 can be compactly imbedded in L^2, there is a strongly convergent subsequence in L^2. Since the convergence is monotone, u_μ is strongly convergent in L^2 as $\mu \to +\infty$. Let $\bar{u}$ be the limit. Then $u_\mu \to \bar{u}$ weakly in H^1_0.

Now we show that $\bar{u}$ solves the variational inequality. If $u \in H^1_0$ satisfies $u \leq b$, then $J(u) = J_\mu(u) \geq J_\mu(u_\mu) \geq J(u_\mu)$. Since J is weakly lower semi-continuous, $J(\bar{u}) \leq \liminf_{\mu \to \infty} J(u_\mu) \leq J(u)$.

It now suffices to show that $\bar{u} \leq b$. To do this, note that $\Phi(\cdot, u_\mu(\cdot))$ is monotone decreasing and always finite. As $\Phi(x, u) = (u - b(x))^2_+$ is bounded below by zero, by the monotone convergence theorem, $\Phi(\cdot, u_\mu(\cdot)) \downarrow \Phi(\cdot, \bar{u}(\cdot))$ in L^1. If $\Phi(\cdot, \bar{u}(\cdot))$ is not zero almost everywhere, then $\|\Phi(\cdot, \bar{u}(\cdot))\|_{L^1} > 0$ and so

$$J_\mu(u_\mu) = J(u_\mu) + \mu \|\Phi(\cdot, u_\mu(\cdot))\|_{L^1} \to +\infty$$

as $\mu \to +\infty$. This contradicts $J_\mu(u_\mu) \leq J(u^*)$. Hence $\Phi(x, \bar{u}(x)) = 0$ for almost all $x \in \Omega$, and so $\bar{u} \leq b$ and $\bar{u}$ solves the variational inequality.

Some further results can be obtained using standard properties of penalty methods. The quantity $J(u_\mu)$ is non-decreasing in μ (see Fletcher[4, **p. 281**]). As $J(u_\mu)$

is bounded above by $J(\bar{u})$, if follows that $J(u_\mu)$ converges as $\mu \to +\infty$. Using the fact that J is weakly LSC,the limit is $\geq J(\bar{u})$. Hence $J(u_\mu) \to J(\bar{u})$. When this is combined with $J_\mu(u_\mu) = J(u_\mu) + \mu\|\Phi(\cdot, u_\mu(\cdot))\|_{L^1} \leq J(\bar{u})$ it is clear that $\mu\|\Phi(\cdot, u_\mu(\cdot))\|_{L^1} \to 0$ as $\mu \to +\infty$.

In this case, if we have an increasing sequence of values of $\mu = \mu_i$, $i = 1, 2, \ldots$, we can generate a sequence u_i by performing one step of Newton's method with $\mu = \mu_i$ starting with u_{i-1}, then $u_1 \geq u_2 \geq \ldots$, and $u_i \geq u_{\mu_i}$. One way of checking that the u_i's do not stray too far from the "central path" is to check that $\mu_i\|\Phi(\cdot, u_i(\cdot))\|_{L^1}$ does indeed go to zero as $i \to \infty$.

5. Solving the discretised equations.

So far, the analysis has all been for the infinite-dimensional problem. Questions remain about how this relates to the actual computations done, which must of necessity be done with discretisations of the true problem. Also, there is the question of how μ becoming "large" affects the ability to solve discretisations of the operator equation

$$-\Delta(\delta u) + \mu\phi_u(x, u)\delta u = f_\mu.$$

Let $h > 0$ be some measure of the coarseness of the discretisation used. Let Δ_h be the discretised form of the operator Δ. It is assumed that $-\Delta_h$ is a symmetric M-matrix. This is true both for standard finite-difference approximations, as well as for piecewise linear finite element approximations. This is important for the monotonicity properties used. Discretising all the unknowns in the above operator equation gives

$$-\Delta_h(\delta u_h) + \mu\phi(x, u)_h\delta u_h = f_{\mu,h}$$

and if $f_h \geq 0$, then $\delta u_h \geq 0$ as N_h-dimensional vectors. Thus, with such approximations, the monotonicity properties of the iterations are retained by the discretised systems.

The next issue is the matter of the speed with which the system of equations can be solved. For systems of this sort, it is usually considered best to use iterative rather than direct methods, although there have been significant improvements in the speed of sparse Cholesky factorisation using techniques such as nested dissection. The problem with using iterative methods, such as conjugate gradients is that as the condition number of the system to be solved increases, the number of iterations needed to solve the system increases. This can be a considerable problem as we wish to take $\mu \to \infty$, and the error in solving the problem is proportional to $1/\mu$.

However, the problem of the number of iterations increasing unboundedly as $\mu \to \infty$ can be solved by using a suitable preconditioner which includes the effect of the large μ.

Such suitable preconditioners are diagonal preconditioners. If the system $(A + \mu D)x = b$ is to be solved for x, and A is symmetric, positive definite and D is a non-negative diagonal matrix, then the preconditioner $\operatorname{diag}(A + \mu D) = \operatorname{diag}(A) + \mu D$ will result in a bound on the pre-conditioned condition number of $A + \mu D$ which is independent of μ for $\mu \geq 0$. In fact, the condition number is no worse than A preconditioned by its diagonal.

The preconditioning matrix is $M = \mathrm{diag}(A) + \mu D$, so the preconditioned matrix whose condition number is needed is $\tilde{A} = M^{-1/2}(A+\mu D)M^{-1/2}$. To obtain bounds for the condition number of $\tilde{A}$, bounds on the extreme eigenvalues can be found:

$$\begin{aligned}
\lambda_{max}(\tilde{A}) &= \max_{v\neq 0} \frac{v^T\tilde{A}v}{v^Tv} \\
&= \max_{w\neq 0} \frac{w^T(A+\mu D)w}{w^T(\mathrm{diag}(A)+\mu D)w} \\
&= \max_{z\neq 0} \frac{z^T \mathrm{diag}(A)^{-1/2}(A+\mu D)\,\mathrm{diag}(A)^{-1/2} z}{z^T(I+\mu\,\mathrm{diag}(A)^{-1}D)z} \\
&= \max_{z\neq 0} \frac{z^T(\hat{A}+\mu\,\mathrm{diag}(A)^{-1}D)z}{z^T(I+\mu\,\mathrm{diag}(A)^{-1}D)z} \\
&\qquad \text{where } \hat{A} = \mathrm{diag}(A)^{-1/2} A\,\mathrm{diag}(A)^{-1/2}.
\end{aligned}$$

Hence

$$\begin{aligned}
\lambda_{max}(\tilde{A}) &\leq \max_{z\neq 0} \frac{z^T(\lambda_{max}(\hat{A})I+\mu\,\mathrm{diag}(A)^{-1}D)z}{z^T(I+\mu\,\mathrm{diag}(A)^{-1}D)z} \\
&\leq \max_{z\neq 0} \frac{\sum_i(\lambda_{max}(\hat{A})+\mu d_{ii}/a_{ii})z_i^2}{\sum_i(1+\mu d_{ii}/a_{ii})z_i^2} \\
&= \max_i \frac{\lambda_{max}(\hat{A})+\mu d_{ii}/a_{ii}}{1+\mu d_{ii}/a_{ii}} \\
&\leq \max(\lambda_{max}(\hat{A}),1) \qquad \text{for all } \mu\geq 0.
\end{aligned}$$

Since every diagonal entry of $\hat{A}$ is one, it follows that $\lambda_{max}(\hat{A}) \geq e_i^T\hat{A}e_i/e_i^Te_i = 1$. Hence $\lambda_{max}(\tilde{A}) \leq \lambda_{max}(\hat{A})$.

Similarly it can be shown that

$$\lambda_{min}(\tilde{A}) \geq \min(\lambda_{min}(\hat{A}),1) \geq \lambda_{min}(\hat{A})$$

for all $\mu \geq 0$. Hence the condition number of the diagonally preconditioned matrix $\tilde{A}$ for all $\mu \geq 0$ is no worse than the condition number of the diagonally preconditioned matrix $\hat{A}$ which is independent of μ. This gives a bound on the number of iterations of preconditioned conjugate gradients needed to achive a particular level of accuracy, which grows, at worst, logarithmically in μ. The growth may be logarithmic as the right-hand side of the equations to be solved grows in magnitude at worst linearly in μ, but the convergence rate is bounded away from zero.

Diagonal preconditioning is not the only, or indeed the best, form of preconditioning. However, it does provide a slowly growing limit on the number of iterations needed as $\mu \to \infty$. The point is perhaps that whatever preconditioner is chosen, it should incorporate the (usually diagonal) part of the discretisation of the $-\Delta + \mu\phi_u(x,u(x))$ operator.

6. Numerical results.

Both exterior penalty and logarithmic barrier algorithms were implemented and numerical results obtained. The test problem used was the problem where Ω is the unit square $[0,1] \times [0,1]$, and $c(x,y) = 10$ and $b(x,y) = \min(x, 1-x, y, 1-y)$. The initial value of μ was taken to be one. A standard five-point finite difference scheme was used for approximating Δ, and the operator $-\Delta + \phi_u(x, u(x))$ is approximated by

$$-\Delta_h + \operatorname{diag}(\phi_u(x_1, u(x_1)), \ldots, \phi_u(x_n, u(x_n))). \tag{6.1}$$

This was used on an $N \times N$ grid (giving $n = N^2$ unknowns). Varying values of N were used to obtain information about the degree of "grid independence" that has actually been achieved. Note that this discretisation preserves the monotonicity of the infinite-dimensional algorithm as $-\Delta_h$ is an M-matrix, and hence, so is (6.1). Thus if the discretised Newton equations are solved accurately enough, the monotonicity of the Newton iterates (and hence of the iterates of the overall algorithm) should be evident.

The iterations were stopped when $\|\delta u_h\|_{L^2} < 10^{-6}$. The computations were done on an IBM RS/6000 computer in the 'C' programming language at the Australian National University's School of Mathematical Sciences.

Fig. 1 shows an illustrative example of the results obtained, which are for $N = 20$.

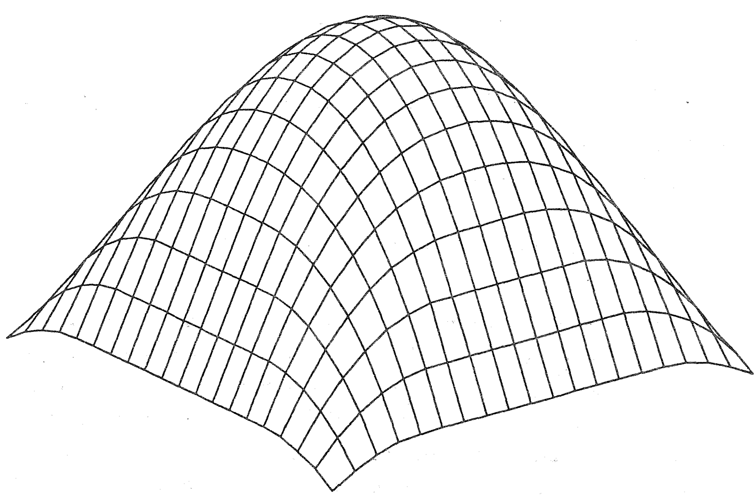

Fig. 1. Solution for $N = 20$

Exterior penalty method.

The exterior penalty function method has been implemented using a number of different preconditioners to solve the linear systems of equations.

The preconditioners used were: the diagonal of the matrix, its incomplete Cholesky factorisation and the SSOR preconditioner $(D+\omega L)D^{-1}(D+\omega L)^T$.

The scheme for updating μ and u is that μ is first increased by a fixed factor of 10^3 until this makes μ exceed 10^7, in which case μ is set to 10^7. Then a single step of Newton's method is used for the resulting value of μ.

Table 1 shows the results using incomplete Cholesky preconditioning for the conjugate gradients algorithm.

N	5	10	20	40	80
# CG iter'ns	27	45	84	227	602
# major cycles	5	5	6	9	12
CPU time (sec)	< 0.1	0.1	0.6	5.8	57.6

Table 1.

Table 2 shows the results using diagonal preconditioning for the conjugate gradients algorithm.

N	5	10	20	40	80
# CG iter'ns	27	64	172	528	1493
# major cycles	5	5	6	9	13
CPU time (sec)	< 0.1	0.1	0.8	7.8	87.6

Table 2.

The results for using an SSOR preconditioner were disappointing; it failed for $N = 10$ by requiring an excessive number of iterations (greater than 10 000) for $\mu = 10^7$.

To indicate the speed of convergence, Table 3 lists how some quantities change with the number of iterations. The values are for using the incomplete Cholesky factorisation preconditioners. As is standard numbers written in the form $x.xx(\pm yy)$ mean $x.xx \times 10^{\pm yy}$. Also, Φ_h is the vector representing $\Phi(x, u(x))$.

The "grid-independence" property of the method is fairly apparent, although it appears that the true infinite-dimensional corrections δu go to zero relatively slowly. This may be due to the nonsmoothness of ϕ. There appears to be a point at which the finite-dimensional approximations suddenly converge quite quickly regardless of the behaviour of the infinite-dimensional problem. This may be due to the method effectively identifying the grid points on the contact surface, which gives rise to an apparent growth in the number of iterations needed to obtain convergence.

This method is competitive with other methods for solving the variational inequality in a grid independent way. However, it is *not* competitive with the fastest methods for solving this class of problems, which appear to be based on a truncated SOR method[3], [5].

Logarithmic barrier methods.

iteration #	1	2	3	4	5	6
μ	1	10^3	10^6	10^7	10^7	10^7
$N=5$						
$\mu\|\Phi_h\|_{L^1}$	0	2.44	6.70	7.9(−5)	7.9(−7)	–
$\|\delta u_h\|_{L^2}$	4.83(−1)	6.32(−2)	3.29(−3)	3.18(−6)	1.08(−9)	–
$N=10$						
$\mu\|\Phi_h\|_{L^1}$	0	2.49	24.4	2.0(−4)	1.9(−6)	–
$\|\delta u_h\|_{L^2}$	4.51(−1)	4.93(−2)	2.63(−2)	1.08(−3)	9.15(−11)	–
$N=20$						
$\mu\|\Phi_h\|_{L^1}$	0	2.41	25.5	3.6(−4)	2.5(−6)	2.4(−6)
$\|\delta u_h\|_{L^2}$	4.32(−1)	4.94(−2)	1.36(−2)	1.07(−2)	1.30(−3)	1.38(−11)
$N=40$						
$\mu\|\Phi_h\|_{L^1}$	0	2.33	27.4	4.4(−4)	3.8(−6)	3.3(−6)
$\|\delta u_h\|_{L^2}$	4.23(−1)	4.59(−2)	1.30(−2)	8.21(−3)	6.05(−3)	3.22(−3)
$N=80$						
$\mu\|\Phi_h\|_{L^1}$	0	2.28	27.1	4.8(−4)	4.5(−6)	4.2(−6)
$\|\delta u_h\|_{L^2}$	4.18(−1)	4.55(−2)	1.11(−2)	4.51(−3)	4.07(−3)	3.52(−3)

Table 3.

A logarithmic barrier method was implemented using a guarded Newton method with Armijo step-length rule: if a given step length results in $\min_{ij}(b_{ij}-u_{ij})\leq 0$ then the step length is halved. If the step length was reduced in any major step, then μ was held fixed until a full step could be taken.

The value of μ was adjusted after every successful full step; μ was then reduced by a fixed ratio until $\mu \leq 10^{-8}$ was reached and thereafter μ was fixed at 10^{-8}. The two ratios used were 10 and 2.

The preconditioner used was the incomplete Cholesky factorisation.

Table 4 shows the performance results for reducing μ be a factor of 10 at each step.

N	5	10	20	40	80
# CG iter'ns	98	185	297	654	1605
# full steps	8	8	8	9	10
# 1/2 steps	3	6	6	6	5
# 1/4 steps	4	5	4	6	11
# $\leq$ 1/8 steps	7	8	9	13	20
# major cycles	22	27	27	34	46
CPU time (sec)	0.1	0.4	2.3	16.8	151.5

Table 4.

Table 5 shows the performance results for reducing μ by a factor of 2 at each step.

The grid-independence of the logarithmic method appears to be somewhat easier to see in the performance results than for the the exterior penalty method. It can be a little harder, however, to directly compare the results for different values of

N	5	10	20	40	80
# CG iter'ns	196	290	513	886	1673
# full steps	22	21	23	23	26
# 1/2 steps	20	21	23	23	19
# 1/4 steps	0	0	0	0	5
# $\leq 1/8$ steps	0	0	0	0	0
# major cycles	42	42	46	48	50
CPU time (sec)	0.2	0.7	3.8	22.8	160.5

Table 5.

N as they usually have different sequences of step lengths, which means that they correspond to different sequences in the infinite-dimensional setting.

Overall, the performance of the exterior penalty method seems to be somewhat better, both in terms of the number of major iterations and in the overall time taken.

REFERENCES

1. E.L. Allgower, K. Böhmer, F.A. Potra and W.C. Rheinboldt, *A mesh-independent principle for operator equations and their discretizations*, SIAM J. Numer. Anal. **23** (1986), 160–169.
2. C. Baiocchi and A. Capelo, *Variational and Quasivariational Inequalities: Applications to Free Boundary Problems*, Wiley, Chichester, New York, 1984.
3. A. Brandt and C.W. Cryer, *Multigrid algorithms for the solution of Linear Complementarity Problems arising from free boundary problems*, SIAM J. Sci. Stat. Comp. **4** (1983), 655–684.
4. R. Fletcher, *Practical Methods of Optimization*, Wiley, Chichester, New York, 1987.
5. R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer–Verlag, New York, 1984, pp. 142–165.
6. H. Minc, *Nonnegative Matrices*, Wiley, New York, London, 1988.
7. J. Schröder, *Operator Inequalities*, Mathematics in Science and Engineering Series #147, Academic Press, New York, London, 1980.
8. S.J. Wright, Private Communcation.

AUSTRALIAN NATIONAL UNIVERSITY, G.P.O. BOX 4, CANBERRA, ACT 2601, AUSTRALIA