Discrete Applied Mathematics 164 (2014) 200–209

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Quasi-optimal degree distribution for a quadratic programming problem arising from the p-version finite element method for a one-dimensional obstacle problem

Matthias Maischak [a,*], Andreas Krebs [b], Ernst P. Stephan [c]

[a] *School of Information Systems, Computing & Mathematics, Brunel University, Uxbridge UB8 3PH, UK*
[b] *Institut für Mathematik, BTU Cottbus, 03013 Cottbus, Germany*
[c] *Institut für Angewandte Mathematik, Universität Hannover, Welfengarten 1, 30167 Hannover, Germany*

ARTICLE INFO

Article history:
Received 15 October 2010
Received in revised form 25 July 2013
Accepted 28 August 2013
Available online 20 September 2013

Keywords:
Quadratic programming
Finite dimensional approximation
Obstacle problem
p-version

ABSTRACT

We present a quadratic programming problem arising from the p-version for a finite element method with an obstacle condition prescribed in Gauss–Lobatto points. We show convergence of the approximate solution to the exact solution in the energy norm. We show an a-priori error estimate and derive an a-posteriori error estimate based on bubble functions which is used in an adaptive p-version. Numerical examples show the superiority of the p-version compared with the h-version.

1. Introduction

We present a quadratic programming problem arising from a p-version finite element scheme for variational inequalities applied to a one-dimensional model problem, where we extend the approach in [10]. We approximate the solution of a continuous minimization problem (1) by a sequence of discrete quadratic programming problems (35), where we optimize the size of the discrete quadratic programming problems, given by the degree vector $\vec{p} \in \mathbb{N}^{|\mathcal{T}_h|}$, $|\mathcal{T}_h|$ is the number of elements in the mesh $\mathcal{T}_h$.

We show asymptotic upper bounds for the minimum of the discrete minimization functional $\underline{J}(\underline{v})$, see (35), depending on the degree vector $\vec{p} \in \mathbb{N}^{|\mathcal{T}_h|}$. First for a uniform degree distribution, see Theorem 2, and second for an adaptively generated degree distribution, see Theorem 3.

Special care has to be taken to ensure that the discretization process is first convergent (see Theorem 1), but also efficient, so that the evaluation of the minimization function is not too expensive. In this scheme the obstacle condition is satisfied in the Gauss–Lobatto points leading to a non-conforming approximation of the convex subset representing the obstacle condition. We prove convergence of the approximate solution to the exact solution of the obstacle problem in the energy norm. For a smooth obstacle our a-priori estimate shows convergence with the rate $\mathcal{O}(p^{-1/2})$ which we believe to be suboptimal. Our proof is based on an abstract error estimate for variational inequalities in [7], which has been applied before only to the h-version with low-order elements (cf. [7,8]). The restriction to the one-dimensional case is for ease

* Corresponding author. Tel.: +44 1895267363.
E-mail addresses: matthias.maischak@brunel.ac.uk, maischak@ifam.uni-hannover.de (M. Maischak), krebs@math.tu-cottbus.de (A. Krebs), stephan@ifam.uni-hannover.de (E.P. Stephan).

0166-218X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dam.2013.08.040

of presentation. The given scheme and the analysis of the a-priori error estimate can be extended straight forwardly to corresponding obstacle problems for second order strongly elliptic or monotone non-linear differential operators in higher dimensions [11], as well as for boundary element methods, see e.g. [13,14].

An alternative to the standard continuous Galerkin method applied here, is the Mortar method, which also leads to consistent discretization schemes for contact problems, see [16].

We present an a-posteriori error estimate based on hierarchical bases for an enriched space of piecewise polynomials of degree $p+1$. Our estimator is based on the stability of the additive Schwarz operator for a two-level decomposition using anti-derivatives of Legendre polynomials. We use this error estimator in an adaptive scheme which increases the polynomial degree of the trial space locally in some subintervals.

Our numerical experiments underline our theory and show better results for the p-version than for the h-version. Contrary to the suboptimal a-priori estimate we obtain for the p-version a higher numerical rate of convergence than for the h-version.

The paper is organized as follows: In Section 2 we show the convergence of the p-version for a one-dimensional model contact problem. Here we make use of the work by Bernardi, Maday about interpolation in Sobolev spaces and by Glowinski, Lions, Trémolières about variational inequalities. The a-posteriori error estimator and the adaptive scheme are given in Section 3. In Section 4 we comment on the implementation issues of the p-version and present numerical experiments. We also compare the p-version with the h-version.

2. Convergence of the *p*-version

Let $V := H_0^1(I) = \{v \in H^1(I), v(\pm 1) = 0\}, I := (-1, 1)$. Let $f \in V' = H^{-1}(I)$. Bilinear form and linear form are defined as

$$a(u, v) := \int_I u'v' \,\mathrm{d}x, \qquad L(v) := \langle f, v\rangle \quad \forall u, v \in V.$$

Furthermore, let $\psi \in H^1(I) \cap C^0(\bar{I})$ be an obstacle function with $\psi(\pm 1) \le 0$, and let K be defined by

$$K := \{v \in H_0^1(I) \mid v \ge \psi \text{ a.e. on } I\}.$$

Then the obstacle problem reads:

$$u = \arg\min_{v\in K} J(v), \qquad J(v) := \frac{1}{2}a(v, v) - L(v), \tag{1}$$

or equivalently, find $u \in K$ such that

$$a(u, v-u) \ge L(v-u) \quad \forall v \in K. \tag{2}$$

Since K is a closed convex nonempty subset of V (cf. [8, Proof of Theorem II.2.1]) it is known from [8, Theorem I.3.1] that there exists a unique solution u of (2).

Next we introduce an approximation of (2) by the p version FEM:

Let $\mathcal{T}_h$ be a finite set of intervals $e \subset \bar{I}$ with $\bar{I} = \bigcup_{e\in\mathcal{T}_h} \bar{e}, \mathring{e}_1 \cap \mathring{e}_2 = \emptyset \forall e_1, e_2 \in \mathcal{T}_h$ with $e_1 \neq e_2$. Set $h := \max_{e\in\mathcal{T}_h} |e|$.

The space V can be approximated by the following p version FE space:

$$V_{\vec{p}} := V_{\vec{p}}^{\mathcal{T}_h} := \{w \in C^0(\bar{I}) \mid w(\pm 1) = 0 \text{ and } w_{|e} \in \mathbb{P}_{p_e}(e),\ \forall\, e \in \mathcal{T}_h\} \tag{3}$$

using the degree vector $\vec{p} \in \mathbb{N}^{|\mathcal{T}_h|}$, with the number of elements $|\mathcal{T}_h$. In the following we investigate first a uniform degree distribution, indicated by V_p, i.e. $p_e = p,\ \forall\, e \in \mathcal{T}_h$.

The continuous minimization problem in general cannot be solved exactly. Therefore it can and has to be replaced by a sequence of discrete minimization problems (arising here from FEM), which depend on the choice of a mesh $\mathcal{T}_h$ and a polynomial degree vector $\vec{p}$. The size of the discrete problem is given by $N = \sum_{e\in\mathcal{T}_h}(p_e + 1)$. Our goal here is to minimize the problem size, by keeping the mesh fixed and choosing an quasi-optimal degree distribution. First, we investigate the approximation quality for a uniform degree distribution, and later we suggest an additive scheme, which allows for a quasi-optimal choice of polynomial degrees.

For the approximation of K we introduce the following notations: On the interval $\bar{I}$ we choose $p+1$ Gauss–Lobatto points, i.e. the points $\xi_j^{p+1}, 0 \le j \le p$, that are the zeros of $(1-\xi^2)L_p'(\xi)$, where L_p denotes the Legendre polynomial of degree p. For these points it is known (cf. [4, Proposition 2.2, (2.3)]) that there exist positive weight factors ρ_j^{p+1} such that

$$\forall \phi \in \mathbb{P}_{2p-1}(\bar{I}) : \sum_{j=0}^{p} \phi(\xi_j)\rho_j^{p+1} = \int_I \phi(\zeta)\,\mathrm{d}\zeta. \tag{4}$$

With the affine linear transformation $Q_e : \bar{I} \to \bar{e} = [a, b]$ given by

$$Q_e(\zeta) := \tfrac{a+b}{2} + \tfrac{b-a}{2}\zeta \tag{5}$$

we define the sets of points

$$G_{e,p} := \{Q_e(\xi_j^{p+1}) \mid 0 \le j \le p\},$$
$$G_p := \bigcup_{e \in \mathcal{T}_h} G_{e,p}.$$

Based on these modified sets of Gauss–Lobatto points the interpolation operator $i_{e,p} : C^0(\bar{e}) \to \mathbb{P}_p(\bar{e})$ is given by

$$(i_{e,p}\psi)(Q_e(\xi_j^{p+1})) = \psi(Q_e(\xi_j^{p+1})), \quad 0 \le j \le p$$

and the global interpolation operator $i_{\mathcal{T}_h,p} : C^0(\bar{I}) \to \mathbb{P}_{\mathcal{T}_h,p}(\bar{I})$ by

$$i_{\mathcal{T}_h,p}\psi = \sum_{e \in \mathcal{T}_h} \chi_e \, i_{e,p}\psi_{|\bar{e}}, \quad \psi \in C^0(\bar{I}). \tag{6}$$

Here, χ_e denotes the characteristic function of e. For the approximation of K we denote the following subset of V_p:

$$K_p := \{w \in V_p \mid w(x_k) \ge \psi(x_k), x_k \in G_p\}. \tag{7}$$

Proposition 1. *K_p is a closed convex subset of V_p.*

Proof. The convexity of K_p is trivial. Let $v_n \to v$ strongly in $H_0^1(I)$, where $v_n \in K_p$ and $v \in H_0^1(I)$. With $v_n(x_k) \ge \psi(x_k)$ $\forall x_k \in G_p$, there follows $v(x_k) \ge \psi(x_k)$ $\forall x_k \in G_p$ and therefore $v \in K_p$.

The discrete problem reads:
Find $u_p \in K_p$ such that

$$a(u_p, v_p - u_p) \ge L(v_p - u_p) \quad \forall v_p \in K_p. \tag{8}$$

From [8, Theorem I.3.1] it follows that (8) has a unique solution. Here, it is worthwhile to note that we have $V_p \subset V$ but we do not have $K_p \subset K$, which means that we have to deal with non-conform approximation sets.

The convergence of the solution u_p of the discrete approximation problem (8) towards the solution u of (2) is stated in the following theorem:

Theorem 1. *Let $\psi \in C^0(\bar{I}) \cap H^1(I)$ and $\psi \le 0$ in a neighborhood of -1 and 1. With the above assumptions on K and K_p, there holds $\lim_{p\to\infty} \|u_p - u\|_{H^1(I)} = 0$ with u_p the solution of (8) and u the solution of (2).*

Proof. The bilinear form $a(\cdot,\cdot)$ is positive definite on V and V_p. Thus, it suffices due to Theorem [8, Theorem I.5.2] to prove the following hypotheses:

H1 If $(v_p)_p$ is such that $v_p \in K_p$, $\forall p$ and converges weakly to v as $p \to \infty$, then $v \in K$.
H2 There exists a dense subset χ of K and a family of mappings $r_p : \chi \to K_p$ such that $\lim_{p\to\infty} r_p v = v$ strongly in V for all $v \in \chi$.

H1 is shown in Lemma 1, H2 in Lemma 2.

Lemma 1. *If the sequence $(v_p)_p$ with $v_p \in K_p$ converges weakly to v for $p \to \infty$ then $v \in K$.*

Proof. Consider $\phi \in C^0(\bar{I})$ with $\phi \ge 0$. For $e \in \mathcal{T}_h$ we approximate ϕ by a combination of Bernstein polynomials on the intervals $e \in \mathcal{T}_h$, i.e.

$$\phi_{e,p}(x) := B_{e,p}\phi(x) := \sum_{k=0}^{p} \binom{p}{k} \left(\frac{x-a}{b-a}\right)^k \left(\frac{b-x}{b-a}\right)^{p-k} \phi_{|e}\left(a + \frac{k}{p}(b-a)\right)$$

and

$$\phi_p := \sum_{e \in \mathcal{T}_h} \chi_e \, \phi_{e,p}(x).$$

Since the Bernstein operators $B_{e,p}$ are monotone we have $\phi_p \ge 0$ and with [5, Theorem 2.3]

$$\lim_{p\to\infty} \|\phi - \phi_p\|_{L^\infty(I)} = 0. \tag{9}$$

For the obstacle function ψ we define the interpolate $\psi_p := i_{\mathcal{T}_h,p}\psi$. By [4, Theorem 4.2] we know

$$\lim_{p\to\infty} \|\psi - \psi_p\|_{L^2(I)} = 0.$$

Now, let the sequence $(v_p)_p$, $v_p \in K_p$ converge weakly to v.

Using (9) and $\phi \in L^\infty(\bar I) = (L^1(\bar I))'$, we obtain that

$$\lim_{p\to\infty} \int_I (v_p - \psi_p)\phi_{p-1}\,dx = \int_I (v-\psi)\phi\,dx\,.$$

With Rellich's embedding theorem (cf. [1, A 5.1]) there follows

$$\lim_{p\to\infty} v_p = v \quad \text{strongly in } L^2(I).$$

Note, that since V is closed and convex it is weakly closed, i.e. $v \in V$ (cf. [1, 5.10]). Thus, it suffices to show that $v \geq \psi$ almost everywhere.

With (4) and the definition of ψ_p we get for all $e \in \mathcal{T}_h$

$$\int_e (v_p - \psi_p)\phi_{p-1}\,dx = \frac{|e|}{2}\sum_{j=0}^{p}[(v_p-\psi_p)\phi_{p-1}](Q_e(\xi_j^{p+1}))\,\rho_j^{p+1} \geq 0. \tag{10}$$

The inequality follows since $\phi_{p-1}(x) \geq 0$ for all $x \in I$ and $(v_p - \psi_p)(x_j) \geq 0$ for all $0 \leq j \leq p$ due to the definition of K_p. Furthermore it is known that the weights ρ_j, $0 \leq j \leq p$, of the Gauss–Lobatto quadrature formula are positive.

Combining (9) and (10) we obtain that for all $\phi \in C^0(\bar I)$ with $\phi \geq 0$

$$\int_e (v-\psi)\phi\,dx \geq 0 \quad \forall e \in \mathcal{T}_h,$$

hence $v \geq \psi$ almost everywhere on I, i.e. $v \in K$.

Lemma 2. *Assuming ψ as in Theorem 1 there exists a dense subset χ of K and a sequence of mappings $r_p : \chi \to K_p$ such that $\lim_{p\to\infty} r_p v = v$ strongly in V for all $v \in \chi$.*

Proof. Consider $\chi := C_0^\infty(I) \cap K$ and $r_p : H_0^1(I) \cap C^0(\bar I) \to V_p$ defined by

$$r_p v := i_{\mathcal{T}_h,p} v. \tag{11}$$

As shown in [4, Theorem 4.5] there exists a constant C independent of v and p such that

$$\|r_p v - v\|_{H^1(I)} \leq Cp^{-1}\|v\|_{H^2(I)} \quad \forall v \in H_0^1(I) \cap H^2(I)$$

and thus for all $v \in \chi$. With (11) it is obvious that $r_p v \in K_p$ for all $v \in \chi$. Thus the assertion of the lemma is fulfilled if $\overline{\chi} = K$. This follows with [8, Lemma II.2.4] if $\psi \leq 0$ in a neighborhood of -1 and 1.

With Theorem 1 the convergence of the p-version is proved without convergence rate.

If we assume higher regularity of the solution u and of the obstacle ψ, i.e. $u \in H_0^2(I)$, $\psi \in H^2(I)$ we obtain the following a priori error estimate which proposes a convergence rate of $\mathcal{O}(p^{-1/2})$. Note, that this assumption is quite natural since a smooth obstacle ψ implies a smooth solution, i.e. $u \in H_0^2(I)$ (cf. [9]). As the proof of Theorem 2 shows, the suboptimal estimate results from our treatment of the term $\|v - u_p\|_{L^2(I)}$ in (14).

Theorem 2. *Let u and u_p be the solutions of (2) and (8), respectively. Furthermore, suppose $u \in H_0^2(I)$, $\psi \in H^2(I)$. Then there exists a constant $C > 0$, independent of u and p, such that*

$$\|u - u_p\|_{H^1(I)} \leq Cp^{-1/2}\left(\|u\|_{H^2(I)} + \|f\|_{L^2(I)} + \|\psi\|_{H^2(I)}\right).$$

Proof. Denote by $A : V \to V'$ the linear map defined, for $u \in V$, by $a(u,v) = \langle Au, v\rangle$ $\forall v \in V$. Let $C_1 > 0$ and $C_2 > 0$ denote the constants of the equivalence inequality

$$C_1\|w\|_{H^1(I)}^2 \leq a(w,w) \leq C_2\|w\|_{H^1(I)}^2 \quad \forall w \in V.$$

Assuming that $f - Au \in L^2(I)$ we apply directly [7, Theorem 1] and obtain

$$\|u-u_p\|_{H^1(I)}^2 \leq \frac{C_2^2}{C_1^2}\|u-v_p\|_{H^1(I)}^2 + \frac{2}{C_1}\|f - Au\|_{L^2(I)}\left(\|u - v_p\|_{L^2(I)} + \|u_p - v\|_{L^2(I)}\right)$$
$$\forall v \in K \text{ and } \forall v_p \in K_p. \tag{12}$$

Let $v_p := i_{\mathcal{T}_h,p}\,u \in K_p$ the interpolate of $u \in H^2(I)$. With [4, Theorems 4.2 and 4.5] we have the following approximation properties:

There exist constants $C_3, C_4 > 0$ independent of u and p such that

$$\begin{aligned}&\|u - v_p\|_{L^2(I)} \leq C_3\,p^{-2}\,\|u\|_{H^2(I)}\\ &\text{and}\quad \|u - v_p\|_{H^1(I)} \leq C_4\,p^{-1}\,\|u\|_{H^2(I)}.\end{aligned} \tag{13}$$

We define the new function $\sup(f_1, f_2)$ by $\sup(f_1, f_2)(x) := \sup(f_1(x), f_2(x))$ a.e. Let v be defined by $v = \sup(u_p - \psi_p, 0) + \psi$ and Φ by $\Phi := \{x \in I : u_p(x) < \psi_p(x)\}$. $u_p - \psi_p|_e$ is a polynomial of degree p on for every element $e \in \mathcal{T}_h$. Therefore $u_p - \psi_p|_e$ can change the sign at most p-times on e. Consequently, Φ is the union of a finite number of open subintervals and $\sup(u_p - \psi_p, 0)$ is continuous and piecewise a polynomial of degree p, i.e. $\sup(u_p - \psi_p, 0) \in H^1$. Then we have $v \in K \cap H^1$.

We can write $v - u_p = \sup(u_p - \psi_p, 0) + \psi - u_p = \psi - \psi_p + \sup(u_p, \psi_p) - u_p$. Due to $u_p \in K_p$ we have $i_{\mathcal{T}_h,p} \sup(u_p, \psi_p) = u_p$ and therefore with [4, Theorem 4.2] that

$$\|v - u_p\|_{L^2(I)} \le \|\psi - \psi_p\|_{L^2(I)} + \|\sup(u_p, \psi_p) - u_p\|_{L^2(I)} \tag{14}$$

$$\le C_3\, p^{-1} \|\psi\|_{H^1(I)} + C_3\, p^{-1} \|\sup(u_p, \psi_p)\|_{H^1(I)}. \tag{15}$$

Using Theorem 1, we know that there exists a C_5 independent of p such that

$$\|\sup(u_p, \psi_p)\|^2_{H^1(I)} = \|u_p\|^2_{H^1(I\setminus\Phi)} + \|\psi_p\|^2_{H^1(\Phi)} \le C_5^2 \left(\|u\|^2_{H^1(I)} + \|\psi\|^2_{H^1(I)} \right). \tag{16}$$

Introducing the coincidence set $\Psi := \{x \in I : u(x) = \psi(x)\}$, we have

$$Au = f \quad \text{on } \Omega \setminus \Psi$$
$$\text{and} \quad Au = A\psi \quad \text{on } \Psi$$

by [9, Theorem II.6.9].

It follows $\|Au\|^2_{L^2(I)} = \|f\|^2_{L^2(I\setminus\Psi)} + \|A\psi\|^2_{L^2(\Psi)} \le \|f\|^2_{L^2(I)} + \|\psi\|^2_{H^2(I)}$, and further

$$\|f - Au\|_{L^2(I)} \le 2\|f\|_{L^2(I)} + \|\psi\|_{H^2(I)}.$$

Combining the error estimates for the interpolation (13), the consistency (16), and for $\|f - Au\|_{L^2(I)}$ with (12), this proves the theorem if we define C as

$$C^2 := \frac{C_2^2 C_4^2}{C_1^2} + \frac{8C_3(1 + C_5)}{C_1}. \tag{17}$$

3. A-posteriori error estimate for the *p*-version

We will extend the space $V_{\vec{p}}$ by bubble functions given on each element in $\mathcal{T}_h$.

Let $L_j(t)$ be the Legendre polynomial of degree j and let $\psi_j(x) := \sqrt{\frac{2j-1}{2}} \int_{-1}^{x} L_{j-1}(t)\, dt$ $(2 \le j)$, $\psi_0(x) := \frac{1-x}{2}$, $\psi_1(x) := \frac{1+x}{2}$. With the affine mapping Q_e as in (5) we define the space

$$\hat{V}_e := \operatorname{span}\{\psi_{e,p_e+1}\} \quad \forall e \in \mathcal{T}_h$$

where $\psi_{e,j} := \psi_j(Q_e^{-1}(x))$.

Hence we obtain the following subspace decomposition

$$V_{\vec{p}+1} := V_{\vec{p}} \oplus \hat{V}_{\vec{p}} \tag{18}$$

where

$$\hat{V}_{\vec{p}} := \sum_{e \in \mathcal{T}_h} \hat{V}_e, \tag{19}$$

and $\vec{p} + 1$ denotes the polynomial degree vector with $(p + 1)_e := p_e + 1$ for all $e \in \mathcal{T}_h$. Due to (7) the convex subset $K_{\vec{p}+1} \subset V_{\vec{p}+1}$ is given.

Let $P_{\vec{p}} : V_{\vec{p}+1} \to V_{\vec{p}}$, $P_{\vec{p},e} : V_{\vec{p}+1} \to \hat{V}_e$ be the Galerkin projections with respect to the bilinear forms $a(\cdot, \cdot)$. For all $u \in V_{\vec{p}+1}$ we define $P_{\vec{p}}$ and $P_{\vec{p},e}$ by

$$a(P_{\vec{p}} u, v) = a(u, v) \quad \forall v \in V_{\vec{p}} \tag{20}$$

$$a(P_{\vec{p},e} u, v) = a(u, v) \quad \forall v \in \hat{V}_e, e \in \mathcal{T}_h. \tag{21}$$

Finally, we define the two-level Schwarz operator

$$P := P_{\vec{p}} + \sum_{e \in \mathcal{T}_h} P_{\vec{p},e}. \tag{22}$$

We use the notation

$$\|\cdot\|_a := a(\cdot, \cdot)^{1/2}.$$

The following lemma states that the condition number of the operator P is independent of $\vec{p}$.

Lemma 3. *There holds*

$$\|v\|_a^2 = \|P_{\vec{p}}v\|_a^2 + \sum_{e\in\mathcal{T}_h} \|P_{\vec{p},e}v\|_a^2 \quad \forall v \in V_{\vec{p}+1} \tag{23}$$

independent of $\vec{p}$ *and* $\mathcal{T}_h$.

Proof. For $e, e' \in \mathcal{T}$ with $e \cap e' = \emptyset$ there holds

$$a(v_e, v_{e'}) = 0 \quad \forall v_e \in \hat{V}_e, v_{e'} \in \hat{V}_{e'} \tag{24}$$

due to the disjointed support of e and e'.

For all $e \in \mathcal{T}_h$ there holds

$$a(v_e, v_p) = 0 \quad \forall v_p \in V_{\vec{p}}, v_e \in \hat{V}_e. \tag{25}$$

This follows, because due to the definition of $V_{\vec{p}}$ the derivative of $v_p|_e(x)$ is a polynomial of degree $p_e - 1$ and the derivative of $v_e(x)$ is the affine image of a Legendre polynomial of degree p_e, therefore the orthogonality property of the Legendre polynomials leads to

$$a(v_e, v_p) = \int_I v_e'(x)v_p'(x)\,dx = \int_e v_e'(x)v_p'|_e(x)\,dx = 0.$$

Let $v = v_p + \sum_e v_e$ with $v_p \in V_p$, $v_e \in \hat{V}_e$ for all $e \in \mathcal{T}_h$. Due to the definition of the Galerkin projections and the orthogonality relations we have

$$\begin{aligned} a(v, v) &= a(v, v_p) + \sum_{e\in\mathcal{T}_h} a(v, v_e) \\ &= a(P_p v, v_p) + \sum_{e\in\mathcal{T}_h} a(P_e v, v_e) \\ &= a(P_p v, v) + \sum_{e\in\mathcal{T}_h} a(P_e v, v) \\ &= a(P_p v, P_p v) + \sum_{e\in\mathcal{T}_h} a(P_e v, P_e v). \end{aligned}$$

Let u be the solution of the variational inequality (2) and let $u_{\vec{p}} \in V_{\vec{p}}$ be the solution of the corresponding discrete problem (8).

We assume the following saturation condition (see, e.g. [2,3]):

Assumption 1. There exist a parameter $0 \le \kappa < 1$ such that for all discrete spaces:

$$\|u - u_{\vec{p}+1}\|_a \le \kappa \|u - u_{\vec{p}}\|_a.$$

Theorem 3. *Assume that Assumption 1 holds. Then, there are constants* $\zeta_1, \zeta_2 > 0$ *such that*

$$\zeta_1 \left(\Theta_{\vec{p}}^2 + \sum_{e\in\mathcal{T}_h} \Theta_{\vec{p},e}^2 \right)^{1/2} \le \|u - u_{\vec{p}}\|_a \le \zeta_2 \left(\Theta_{\vec{p}}^2 + \sum_{e\in\mathcal{T}_h} \Theta_{\vec{p},e}^2 \right)^{1/2} \tag{26}$$

where

$$\Theta_{\vec{p}} := \|P_{\vec{p}} e_{\vec{p}+1}\|_a \tag{27}$$

$$\Theta_{\vec{p},e} := \|P_e e_{\vec{p}+1}\|_a = \frac{|a(e_{\vec{p}+1}, \psi_{e,p_e+1})|}{\|\psi_{e,p_e+1}\|_a} \tag{28}$$

and $e_{\vec{p}+1} \in K_{\vec{p}+1} - u_{\vec{p}}$ *is the solution of the variational inequality*

$$a(e_{\vec{p}+1}, v - e_{\vec{p}+1}) \ge L(v - e_{\vec{p}+1}) - a(u_{\vec{p}}; v - e_{\vec{p}+1}) \quad \forall v \in K_{\vec{p}+1} - u_{\vec{p}} \tag{29}$$

with $K_{\vec{p}+1} - u_{\vec{p}} := \{v - u_{\vec{p}} \mid v \in K_{\vec{p}+1}\}$.

Proof. The defect $e_{\vec{p}+1} := u_{\vec{p}+1} - u_{\vec{p}} \in K_{\vec{p}+1} - u_{\vec{p}}$ of the solution to

$$a(u_{\vec{p}+1}; v - u_{\vec{p}+1}) \geq L(v - u_{\vec{p}+1}) \quad \forall v \in K_{\vec{p}+1} \tag{30}$$

is given by the solution of

$$a(e_{\vec{p}+1}; v - e_{\vec{p}+1}) \geq L(v - e_{\vec{p}+1}) - a(u_{\vec{p}}; v - e_{\vec{p}+1}) \quad \forall v \in K_{\vec{p}+1} - u_{\vec{p}}. \tag{31}$$

(31) is solvable analogously to (8) with the shifted obstacle function $\psi - u_{\vec{p}}$.

We can apply Lemma 3 to obtain

$$\|e_{\vec{p}+1}\|_a^2 = \|P_{\vec{p}} e_{\vec{p}+1}\|_a^2 + \sum_{e \in \mathcal{T}_h} \|P_e e_{\vec{p}+1}\|_a^2. \tag{32}$$

Applying the saturation Assumption 1 we obtain

$$(1 - \kappa)\|u - u_{\vec{p}}\|_a \leq \|u_{\vec{p}+1} - u_{\vec{p}}\|_a = \|e_{\vec{p}+1}\|_a \leq (1 + \kappa)\|u - u_{\vec{p}}\|_a.$$

Combining this with (32) we obtain the assertion of this theorem with the constants

$$\zeta_1 = \frac{1}{1 + \kappa} \quad \text{and} \quad \zeta_2 = \frac{1}{1 - \kappa}. \tag{33}$$

The coarse grid error estimator $\Theta_{\vec{p}}^2 = \|P_{\vec{p}} e_{\vec{p}+1}\|_a^2$ has to be computed by solving explicitly the linear system

$$a(P_p e_{p+1}, v) = a(e_{p+1}, v) \quad \forall v \in V_{\vec{p}}$$

to obtain $P_p e_{p+1} \in V_{\vec{p}}$.

The fine grid error estimators $\Theta_{\vec{p},e} \in \hat{V}_e$ have to be computed by solving the one dimensional linear systems (21), i.e. for all $e \in \mathcal{T}_h$ we have to solve

$$a(P_{\vec{p},e} e_{\vec{p}+1}, v) = a(e_{\vec{p}+1}, v) \quad \forall v \in \hat{V}_e.$$

Due to $P_{\vec{p},e} e_{\vec{p}+1} \in \hat{V}_e$ we have $P_{\vec{p},e} e_{\vec{p}+1} = \mu_e \psi_{e,p_e+1}$ and therefore

$$a(P_{\vec{p},e} e_{\vec{p}+1}, \psi_{e,p_e+1}) = \mu_e a(\psi_{e,p_e+1}, \psi_{e,p_e+1}) = a(e_{\vec{p}+1}, \psi_{e,p_e+1}),$$

i.e. $\mu_e = \frac{a(e_{\vec{p}+1}, \psi_{e,p_e+1})}{a(\psi_{e,p_e+1}, \psi_{e,p_e+1})}$ and

$$\Theta_{\vec{p}}^2 = a(P_{\vec{p}} e_{\vec{p}+1}, P_{\vec{p}} e_{\vec{p}+1}) = a(P_{\vec{p}} e_{\vec{p}+1}, e_{\vec{p}+1}) = \frac{a^2(e_{\vec{p}+1}, \psi_{e,p_e+1})}{a(\psi_{e,p_e+1}, \psi_{e,p_e+1})}.$$

Remark 1. We have to compute the solution of (29) explicitly, because (29) is a variational inequality. The solutions of variational inequalities do not have the orthogonality property of the Galerkin solution of a usual variational formulation. In case of a standard Galerkin solution the coarse grid error estimator is vanishing due to the Galerkin orthogonality.

Adaptive algorithm for the p-version: In this section we formulate an adaptive algorithm which uses the error indicators from Theorem 3 to generate a sequence of locally enriched spaces.

We estimate the global error by

$$\eta_{\vec{p}} := \left(\Theta_{\vec{p}}^2 + \sum_{e \in \mathcal{T}_h} \Theta_{\vec{p},e}^2 \right)^{1/2}. \tag{34}$$

Algorithm 1. Let the parameters $\epsilon > 0$, $0 < \delta < 1$ and an subdivision $\mathcal{T}_h$ of I and an initial polynomial degree vector $\vec{p}_0$ be given. With $V_{(\vec{p})_0}$ we denote the initial test and trial space.

For $k = 0, 1, 2, \ldots$

1. Compute the solution $u_{(\vec{p})_k} \in K_{(\vec{p})_k} \subset V_{(\vec{p})_k}$ of (8).
2. Solve the defect problem (29) for $e_{\vec{p}+1}$.
3. Compute the error indicator $\Theta_{\vec{p}}$.
4. Compute the local error indicator $\Theta_{\vec{p},e}$ for each $e \in \mathcal{T}_h$.
5. Compute the global error estimate $\eta_{\vec{p}}$. Stop if $\eta_{\vec{p}} < \epsilon$.
6. Determine $\Theta_{\vec{p},e'}$ such that $\text{card}(\{e \in \mathcal{T}_h : \Theta_{\vec{p},e} < \Theta_{\vec{p},e'}\}) = \lfloor \delta \text{card}(\mathcal{T}_h) \rfloor$.
7. p-adaption step. If $\Theta_{\vec{p},e} \geq \Theta_{\vec{p},e'}$, increase the polynomial degree on $e \in \mathcal{T}_h$ by 1. This defines an enlarged space $V_{(\vec{p})_{k+1}} \supset V_{(\vec{p})_k}$. Goto 1.

Table 1
Convergence of the h-version.

i	N_i	$\|u-u_i\|_{H^1(I)}$	α_i	C_i	#It	CPU (s)
1	15	$7.22 \cdot 10^{-2}$	–	–	17	0.00
2	31	$3.60 \cdot 10^{-2}$	0.96	0.96	33	0.00
3	63	$1.80 \cdot 10^{-2}$	0.98	1.04	85	0.01
4	127	$9.02 \cdot 10^{-3}$	0.99	1.09	201	0.02
5	255	$4.51 \cdot 10^{-3}$	0.99	1.11	472	0.01
6	511	$2.25 \cdot 10^{-3}$	1.00	1.13	835	0.02
7	1 023	$1.12 \cdot 10^{-3}$	1.00	1.14	4 401	0.16
8	2 047	$5.63 \cdot 10^{-4}$	1.00	1.15	8 219	0.53
9	4 095	$2.82 \cdot 10^{-4}$	1.00	1.15	23 348	2.93
10	8 191	$1.41 \cdot 10^{-4}$	1.00	1.15	59 329	14.14
11	16 383	$7.05 \cdot 10^{-5}$	1.00	1.15	141 182	65.25

4. Implementation and numerical experiments

For the implementation we define a basis B of V_p by the Lagrange interpolation polynomials $\lambda_{e,j}$ on the intervals $e \in \mathcal{T}_h$ corresponding to the Gauss–Lobatto points $Q_e(\xi_j^{p+1})$, $0 \le j \le p$. Defining

$$\tilde{\lambda}_i^p(\xi) := \prod_{0\le k\le p,\ k\ne i} \frac{\xi - \xi_k}{\xi_i - \xi_k} \quad \forall 0 \le k \le p$$

where ξ_k denote the Gauss–Lobatto points on I we have $\tilde{\lambda}_i^p(\xi_j) = \delta_{ij}$. Thus with $\lambda_{e,i}(x) = \tilde{\lambda}_i^p(Q_e^{-1}(x))$, $\forall x \in e$, and $\lambda_{e,i}(x) \equiv 0$, $\forall x \in \bar{I}\backslash\bar{e}$, we can rewrite K_p of (7):

$$K_p = \left\{ \sum_{e\in\mathcal{T}_h} \sum_{0\le i\le p} w_{e,i}\lambda_{e,i} \mid w_{e,i} \ge \psi(Q_e(\xi_i)),\ \forall e \in \mathcal{T}_h,\ 0 \le i \le p \right\}.$$

Introducing a global counting for the pairs (e, i) such that $B = \{\lambda_j \mid 1 \le k \le N\}$, $N := \dim B$ and $\psi_j := \psi(Q_e(\xi_i))$ the discrete problem (8) can be rewritten as a matrix inequality:

Find $\underline{u} \in \underline{K}_p$ such that

$$(\underline{v}^T - \underline{u}^T)\,\underline{\underline{A}}\,\underline{u} \ge (\underline{v}^T - \underline{u}^T)\,\underline{f}, \quad \forall \underline{v} \in \underline{K}_p$$

where $\underline{K}_p, \underline{\underline{A}}, \underline{f}$ are defined by $\underline{K}_p := \{\underline{w} \in \mathbb{R}^N \mid w_j \ge \psi_j,\ 1 \le j \le N\}$, $\underline{f} := (L(\lambda_j))_{1\le j\le N}$, and $\underline{\underline{A}} := (a(\lambda_j, \lambda_i))_{1\le i\le N,\,1\le j\le N}$. This is equivalent to the quadratic programming problem

$$\underline{u} = \arg\min_{\underline{v}\in\underline{K}_p} \underline{J}(\underline{v}), \qquad \underline{J}(\underline{v}) := \frac{1}{2}\underline{v}^T\,\underline{\underline{A}}\,\underline{v} - \underline{v}^T\,\underline{f}. \tag{35}$$

This problem of minimizing a convex quadratic form subject to upper or lower bounds on the variables can be solved by relaxation methods (cf. [8, Chapter V]) or a generalized conjugate gradient algorithm (cf. [15]), known as Polyak algorithm. A more modern and also more efficient alternative is the MPRGP-algorithm found in [6].

The matrix inequalities of the following example were solved by the Polyak algorithm. As initial values for the iteration process we prolongated the known solution $u_p \in V_p$ into $u_{p'} \in V_{p'}$, $p' > p$, via basis transformation. The prolongation decreases the time for computation significantly and leads to stable solutions $u_{p'}$ in spite of the high condition number of the matrix A.

Example. With $\psi = |x| - 1$, $x \in I$, $f \equiv 2$ the exact solution of (2) is given by (cf. [8, Section II 3.3.2])

$$u = \begin{cases} -x - 1 & \text{for } x \le -\frac{1}{2} \\ x - 1 & \text{for } x \ge \frac{1}{2} \\ x^2 - \frac{3}{4} & \text{else.} \end{cases}$$

First we compute the discrete solution of (8) via the h-version. In Table 1 we list the computed values for the experimental convergence rate

$$\alpha_i^1 := -\frac{\log\big(\|u - u_i\|_{H^1(I)} / \|u - u_{i-1}\|_{H^1(I)}\big)}{\log\big(N_i / N_{i-1}\big)} \quad \text{and} \quad C_i^1 := N_i^{-\alpha_i} \cdot \|u - u_i\|_{H^1(I)};$$

Table 2
Convergence of the p-version, $h = 2/5$.

p	N_p	$\|u - u_p\|_{H_1(I)}$	α_i	p	N_p	$\|u - u_p\|_{H_1(I)}$	α_i
2	9	$5.2701 \cdot 10^{-2}$		4	19	$8.1986 \cdot 10^{-3}$	
5	24	$1.0758 \cdot 10^{-2}$	−1.734	7	34	$3.8914 \cdot 10^{-3}$	−1.331
8	39	$5.2139 \cdot 10^{-3}$	−1.541	10	49	$2.3912 \cdot 10^{-3}$	−1.365
11	54	$3.2205 \cdot 10^{-3}$	−1.513	13	64	$1.6640 \cdot 10^{-3}$	−1.381
14	69	$2.2411 \cdot 10^{-3}$	−1.503	16	79	$1.2474 \cdot 10^{-3}$	−1.388
17	74	$1.6752 \cdot 10^{-3}$	−1.499	19	94	$9.8273 \cdot 10^{-4}$	−1.387
20	99	$1.3137 \cdot 10^{-3}$	−1.496	22	109	$8.0246 \cdot 10^{-4}$	−1.382
23	114	$1.0662 \cdot 10^{-3}$	−1.493	25	124	$6.7324 \cdot 10^{-4}$	−1.373
26	129	$8.8824 \cdot 10^{-4}$	−1.490	28	139	$5.7694 \cdot 10^{-4}$	−1.362
29	144	$7.5517 \cdot 10^{-4}$	−1.486	31	154	$5.0294 \cdot 10^{-4}$	−1.348
32	159	$6.5270 \cdot 10^{-4}$	−1.481	34	169	$4.4463 \cdot 10^{-4}$	−1.333
35	174	$5.7187 \cdot 10^{-4}$	−1.475	37	184	$3.9774 \cdot 10^{-4}$	−1.318
38	189	$5.0685 \cdot 10^{-4}$	−1.468	40	199	$3.5936 \cdot 10^{-4}$	−1.301
41	204	$4.5370 \cdot 10^{-4}$	−1.458	43	214	$3.2747 \cdot 10^{-4}$	−1.284
44	219	$4.0966 \cdot 10^{-4}$	−1.446	46	229	$3.0063 \cdot 10^{-4}$	−1.268
47	234	$3.7276 \cdot 10^{-4}$	−1.431	49	244	$2.7777 \cdot 10^{-4}$	−1.251

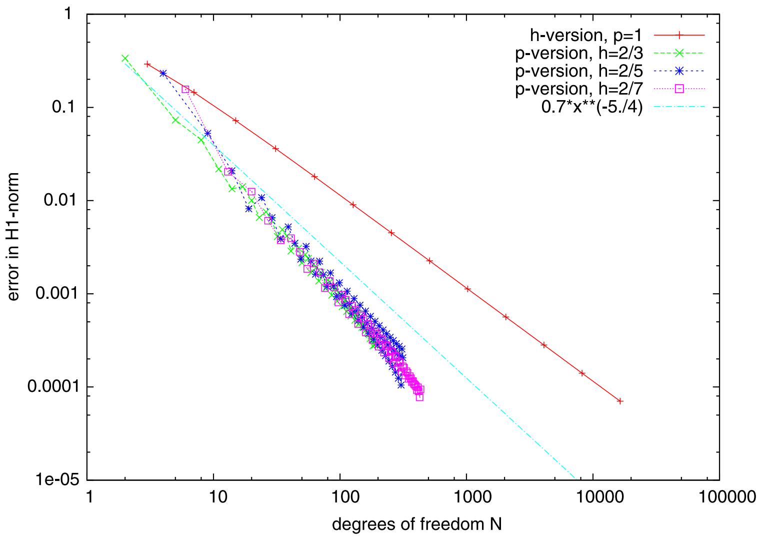

Fig. 1. Convergence of the h-version and of the p-version for different mesh parameters h.

we also give the total iteration number '#It' for the Polyak algorithm and the cpu-time 'CPU(s)' for the solver. The stopping criterion for the iterative solver was $\|x^{k+1} - x^k\|_2 \le 10^{-10}\|x_k\|_2$.

For the p-version we define $\mathcal{T}_h := \{\,]-1 + (i-1)h, -1 + ih[, \mid 1 \le i \le 2/h\}$. The errors and the experimental convergence rate for $h := 2/5$ are listed in Table 2, in the left part p runs from 2 up to 47 with step width 3, and in the right part p runs from 4 up to 49 with step width 3. Fig. 1 shows the results for $h := 2/3,\ 2/5,\ 2/7$ with p up to 63. The oscillating errors of Fig. 1 are due to the non-conformities $K_{p+1} \not\subseteq K_p \not\subseteq K$. Thus, we do not obtain a strict convergence in the sense of $\|u - u_{p+1}\|_{H^1(I)} < \|u - u_p\|_{H^1(I)}$, but we observe a 3-step structure, see Table 2, where we can compute a numerical convergence rate. Nevertheless, Table 2 and Fig. 1 confirm

$$\|u - u_{N(h,p)}\|_{H^1(I)} < 0.7N(h,p)^{-5/4} \le \mathcal{O}(p^{-5/4})$$

for different mesh parameters h and a better convergence rate than in the case of the h-version concerning the number of unknowns $N(h, 1)$ and $N(h, p)$, respectively.

Due to our discrete formulation our basis functions are necessarily Lagrange polynomials in Gauss–Lobatto points. In Table 3 we give the extreme eigenvalues and condition numbers for the resulting stiffness matrix of the p-version with $h = 2/5$. The condition number grows $\kappa(\underline{A}) \sim \mathcal{O}(p^{5/2})$.

The computations have been performed with the research framework *Maiprogs* [12] using a Intel i7-3820QM machine with 2.70 GHz, 4 cores and 16 GB main memory.

Table 3
Eigenvalues λ_{min}, λ_{max} and condition number κ of the p-version, $h = 2/5$.

p	N_p	λ_{min}	λ_{max}	κ	p	N_p	λ_{min}	λ_{max}	κ
2	9	0.49242	25.85	$5.250 \cdot 10^1$	4	19	0.24639	70.58	$2.864 \cdot 10^2$
5	24	0.19710	97.54	$4.948 \cdot 10^2$	7	34	0.14076	173.4	$1.232 \cdot 10^3$
8	39	0.12315	221.4	$1.798 \cdot 10^3$	10	49	0.09851	333.4	$3.384 \cdot 10^3$
11	54	0.08955	398.5	$4.450 \cdot 10^3$	13	64	0.07577	546.6	$7.214 \cdot 10^3$
14	69	0.07036	629.6	$8.949 \cdot 10^3$	16	79	0.06156	813.5	$1.322 \cdot 10^4$
17	84	0.05794	914.4	$1.578 \cdot 10^4$	19	94	0.05183	1134.0	$2.188 \cdot 10^4$
20	99	0.04924	1252.0	$2.544 \cdot 10^4$	22	109	0.04476	1507.0	$3.368 \cdot 10^4$
23	114	0.04282	1644.0	$3.840 \cdot 10^4$	25	124	0.03939	1935.0	$4.913 \cdot 10^4$
26	129	0.03788	2089.0	$5.517 \cdot 10^4$	28	139	0.03517	2416.0	$6.869 \cdot 10^4$
29	144	0.03396	2588.0	$7.622 \cdot 10^4$	31	154	0.03176	2950.0	$9.287 \cdot 10^4$
32	159	0.03077	3140.0	$1.020 \cdot 10^5$	34	169	0.02896	3538.0	$1.222 \cdot 10^5$
35	174	0.02813	3746.0	$1.331 \cdot 10^5$	37	184	0.02661	4179.0	$1.570 \cdot 10^5$
38	189	0.02591	4404.0	$1.700 \cdot 10^5$	40	199	0.02462	4873.0	$1.980 \cdot 10^5$
41	204	0.02401	5117.0	$2.130 \cdot 10^5$	43	214	0.02290	5622.0	$2.455 \cdot 10^5$
44	219	0.02238	5883.0	$2.629 \cdot 10^5$	46	229	0.02140	6423.0	$3.000 \cdot 10^5$
47	234	0.02095	6702.0	$3.199 \cdot 10^5$	49	244	0.02009	7278.0	$3.621 \cdot 10^5$

References

[1] H.W. Alt, Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung (Linear Functional Analysis. An Application Oriented Introduction), in: überarb. und erw. Auflage., vollst. 3, Springer, Springer-Lehrbuch, Berlin, 1999. xiii, 415 S.

[2] R.E. Bank, Hierarchical bases and the finite element method, Acta Numer. (1996) 1–43.

[3] R. Bank, R. Smith, A posteriori error estimates based on hierarchical bases, SIAM J. Numer. Anal. 30 (1993) 921–935.

[4] C. Bernardi, Y. Maday, Polynomial interpolation results in Sobolev spaces, J. Comput. Appl. Math. 43 (1992) 53–80.

[5] R.A. DeVore, G.G. Lorentz, Constructive Approximation: Polynomials and Splines Approximation, in: Grundlehren der Mathematischen Wissenschaften, vol. 303, Springer-Verlag, Berlin, 1993.

[6] Z. Dostál, J. Schöberl, Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination, Comput. Optim. Appl. 30 (1) (2005) 23–44.

[7] R.S. Falk, Error estimates for the approximation of a class of variational inequalities, Math. Comp. 28 (1974) 963–971.

[8] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, in: Springer Series in Computational Physics, Springer-Verlag, Berlin, 1984.

[9] D. Kinderlehrer, G. Stampacchia, An Introduction to Variational Inequalities and Their Applications, in: Pure and Applied Mathematics, vol. 88, Academic Press, New York etc., 1980, p. 313. XIV.

[10] A. Krebs, M. Maischak, E.P. Stephan, A quadratic programming problem arising from the *p*-version finite element method for a one-dimensional obstacle problem, Electron. Notes Discrete Math. 36 (2010) 923–930.

[11] A. Krebs, E.P. Stephan, A *p*-version finite element method for nonlinear elliptic variational inequalities in 2d, Numer. Math. 105 (3) (2007) 457–480.

[12] M. Maischak, The software package maiprogs, http://www.ifam.uni-hannover.de/~maiprogs.

[13] M. Maischak, E.P. Stephan, Adaptive *hp* versions of BEM for Signorini problems, Appl. Numer. Math. 54 (3–4) (2005) 425–449.

[14] M. Maischak, E.P. Stephan, Adaptive *hp*-versions of boundary element methods for elastic contact problems, Comput. Mech. 39 (5) (2007) 597–607.

[15] D.P. O'Leary, A generalized conjugate gradient algorithm for solving a class of quadratic programming problems, Linear Algebra Appl. 34 (1980) 371–399.

[16] B. Wohlmuth, Variationally consistent discretization scheme and numerical algorithms for contact problems, Acta Numer. (2011) 569–734.