

Finite element methods for elastic contact: penalty and Nitsche

Tom Gustafsson and Rolf Stenberg¹

Summary We consider two methods for treating elastic contact problems with the finite element method; the penalty method and Nitsche’s method. For the penalty method we discuss how the penalty parameter should be chosen. Both the theoretical analysis and numerical examples show that an optimal convergence rate cannot be achieved. The method is contrasted to that of Nitsche’s method which is optimally convergent. We also give the derivation of Nitsche’s method by a very simple consistency correction of the penalty method.

Keywords: FEM, penalty method, Nitsche’s method, contact problems

Received: 24 November 2024. Accepted: 8 May 2025. Published online: 23 June 2025.

Introduction

The paper *Variational methods for the solution of problems of equilibrium and vibrations* [3], presented by Richard Courant at the American Mathematics Society meeting in Washington D.C. in 1941, is widely considered as a starting point in two fields of applied mathematics, finite element methods and the use of the penalty method in optimization. For the case of optimization this appears to be the case, see, e.g., the classical books [5, 12].

In the field of finite elements, the continuous piecewise linear basis functions are often referred to the ”Courant triangle” or ”Courant hat function”. The paper by Courant came out in print in 1943, and to celebrate the 50th anniversary, a conference was organized in Jyväskylä [9]. Ivo Babuška gave an invited talk, in which he outlined the history of the finite element method. Surprisingly, he discovered that none of the earliest paper on the mathematical foundations of the FE cite Courants paper. Furthermore, he pointed out that piecewise linear approximation method was not presented at the conference, it was introduced as an appendix in the final paper.

The same comment is given by Gilbert Strang in the first textbook (Strang & Fix [16]) on the mathematics of finite elements: *the penalty method was the main theme of Courant’s remarkable lecture; the finite element method was an afterthought.*

¹Corresponding author: rolf.stenberg@aalto.fi

Today, the penalty method is widely used in finite elements, especially in applied engineering. From a mathematical point of view, the method is non-conforming and the central question is how to choose the penalty parameter. The engineering literature is rather vague on this issue, cf. [10, 18]. The same applies for some commercial codes that we checked, with the COMSOL as an exception.

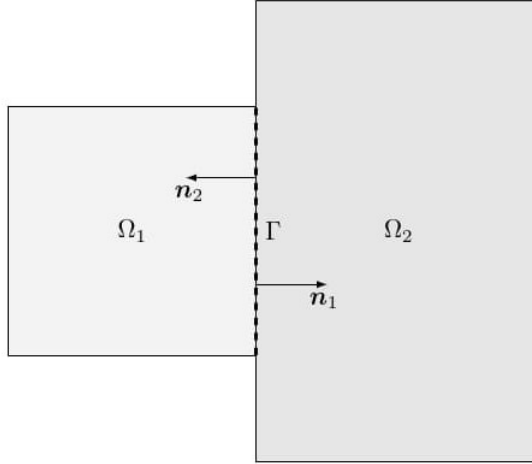
Below we will discuss this matter and we will also recall Nitsche's method which gives a simple fix to the method.

The penalty method

All the questions we are going to address can be answered by considering "mortaring" of the Poisson equation. To consider elasticity only adds more notation. To consider a contact problem where the contact zone is unknown leads to technicalities involving variational inequalities, cf. [7]. By treating both of these would only blur the message of this paper.

Hence, we consider the Poisson problem in a domain $\Omega \subset \mathbb{R}^n$, that we split in two $\Omega = \Omega_1 \cup \Omega_2$, with the continuity conditions at the common boundary $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$: find $u = (u_1, u_2)$ such that

$$\begin{aligned} -\Delta u_i &= f && \text{in } \Omega_i, \\ u_1 - u_2 &= 0 && \text{on } \Gamma, \\ \frac{\partial u_1}{\partial n_1} + \frac{\partial u_2}{\partial n_2} &= 0 && \text{on } \Gamma, \\ u_i &= 0 && \text{on } \partial\Omega_i \setminus \Gamma. \end{aligned} \tag{1}$$



Clearly U , with $U|_{\Omega_i} = u_i$, $i = 1, 2$, solves the equation

$$-\Delta U = f \text{ in } \Omega, \quad U = 0 \text{ on } \partial\Omega. \tag{2}$$

The continuous penalized problem is: Find $u_\varepsilon = (u_1^\varepsilon, u_2^\varepsilon)$ minimizing

$$E(v) = \sum_{i=1}^2 \left(\frac{1}{2} \int_{\Omega_i} |\nabla v_i|^2 dx - \int_{\Omega_i} f v_i dx \right) + \int_{\Gamma} \frac{1}{2\varepsilon} (v_1 - v_2)^2 ds. \tag{3}$$

It is clear that a necessary condition in the limit $\varepsilon \rightarrow 0$ is that $u_1^0 = u_2^0$ on Γ , with $\lim_{\varepsilon \rightarrow 0} u_i^\varepsilon = u_i^0 = u_i$, $i = 1, 2$.

The physical interpretation of non-conformity of the method is that the problem is changed to that where a distributed spring is introduced on Γ , giving the interface condition

$$\frac{\partial u_i^\varepsilon}{\partial n_i} + \varepsilon^{-1}(u_i^\varepsilon - u_j^\varepsilon) = 0, \quad i \neq j.$$

In the FEM, we perform the minimization in subspace V_h of piecewise polynomials of degree p .

The Finite Element Method: Find $u_h^\varepsilon = (u_{1,h}^\varepsilon, u_{2,h}^\varepsilon) \in V_h = V_{1,h} \times V_{2,h}$ that minimizes the energy $E(v)$ in the subspace V_h .

The *accuracy* of the FE solution is most naturally measured in *energy norm*:

$$\|v\|_E^2 = \sum_{i=1}^2 \int_{\Omega_i} |\nabla v_i|^2 dx + \int_{\Gamma} \frac{1}{\varepsilon} (v_1 - v_2)^2 ds. \quad (4)$$

The basic mathematical result is that the *FE solution is the best approximation in the energy norm*, viz.

$$\|u_\varepsilon - u_h^\varepsilon\|_E \lesssim \min_{v_h \in V_h} \left\{ \sum_{i=1}^2 \|\nabla(u_i^\varepsilon - v_{i,h})\|_{0,\Omega_i} + \varepsilon^{-1/2} \|u_i^\varepsilon - v_{i,h}\|_{0,\Gamma} \right\}. \quad (5)$$

For notational simplicity we now assume that ε is a constant along Γ .

Note that for FEM the exact solution is that of the penalized exact solution u_ε . To the total error we should thus add the error due to lack of consistency or, alternatively stated, modelling. It is not too difficult to show that this is

$$\|u - u_\varepsilon\|_E \lesssim \varepsilon^{1/2} \left\| \frac{\partial u}{\partial n} \right\|_{0,\Gamma}, \quad (6)$$

where $\lesssim \dots$ stands for: there exists a positive constant C , such that $\leq C \dots$.

By the triangle inequality we have

$$\|u - u_h^\varepsilon\|_E \leq \|u - u_\varepsilon\|_E + \|u_\varepsilon - u_h^\varepsilon\|_E. \quad (7)$$

As usual, let h denote the global mesh length.

With the assumed smoothness of the exact penalized solution $u_\varepsilon \in H^s(\Omega)$, with $1 \leq s \leq p+1$, interpolation theory yields

$$\min_{v_h \in V_h} \left\{ \sum_{i=1}^2 \|\nabla(u_i^\varepsilon - v_{i,h})\|_{0,\Omega_i} + \varepsilon^{-1/2} \|u_i^\varepsilon - v_{i,h}\|_{0,\Gamma} \right\} \lesssim (h^{s-1} + \varepsilon^{-1/2} h^{s-1/2}) \|u_\varepsilon\|_s. \quad (8)$$

(This estimate was first given by Babuška [1].) By choosing $\varepsilon \approx h$, we get the optimal convergence rate

$$\|u_\varepsilon - u_h^\varepsilon\|_E \lesssim h^s \|u_\varepsilon\|_s. \quad (9)$$

From (6) and (7) we thus get

$$\|u - u_h^\varepsilon\|_E \lesssim h^{1/2} \left\| \frac{\partial u}{\partial n} \right\|_{0,\Gamma} + h^s \|u_\varepsilon\|_s. \quad (10)$$

From above we conclude that the modelling error is dominating, except for a very singular solution.

The modelling error can of course be made smaller by decreasing ε . However, (8) shows that for a given smoothness, the interpolation error will increase. The condition number also increases from $\mathcal{O}(h^{-2})$, which is the one expected for a second order elliptic equation.

In conclusion: *with the penalty method we will never obtain a method with optimal convergence rate.*

Next, let us turn to the *a posteriori* error analysis. In addition to the normal terms, and edge $E \subset \Gamma$ yields the term [8]

$$\eta_E^2 = h_E + h_E \left\| \frac{\partial u_{i,h}^\varepsilon}{\partial n_i} + \varepsilon^{-1} (u_{i,h}^\varepsilon - u_{j,h}^\varepsilon) \right\|_{0,E}^2, \quad j \neq i. \quad (11)$$

In [8] we also showed that the second term above is efficient, i.e. bounded by the real error if and only if $\varepsilon \approx h$, giving an additional justification for this choice.

In the next section we will show that there is a very simple correction that leads to an optimal method, i.e. Nitsche's method.

Nitsche's method [11]

Let us recall that the weak form obtained by minimizing (3) is: find $u_\varepsilon = (u_1^\varepsilon, u_2^\varepsilon)$ such that

$$\sum_{i=1}^2 \left(\int_{\Omega_i} \nabla u_i^\varepsilon \cdot \nabla v_i \, dx + \int_{\Gamma} \frac{1}{\varepsilon} (u_i^\varepsilon - u_j) v_i \, ds \right) = \sum_{i=1}^2 \int_{\Omega_i} f v_i \, dx \quad (12)$$

for all test functions $v = (v_1, v_2)$.

For the solution $u = (u_1, u_2)$ of the exact problem (1) it holds

$$u_i = u_j \quad \text{on } \Gamma, \quad (13)$$

and hence

$$\int_{\Gamma} \frac{1}{\varepsilon} (u_i - u_j) v_i \, ds = 0. \quad (14)$$

Further, multiplying the differential equation in (1), integrating over Ω_i , and integrating by parts yields

$$\int_{\Omega_i} \nabla u_i \cdot \nabla v_i \, dx - \int_{\Gamma} \frac{\partial u_i}{\partial n_i} v_i \, ds = \int_{\Omega_i} f v_i \, dx. \quad (15)$$

These two relations show that the exact solution satisfies

$$\begin{aligned} & \sum_{i=1}^2 \left(\int_{\Omega_i} \nabla u_i \cdot \nabla v_i \, dx + \int_{\Gamma} \frac{1}{\varepsilon} (u_i - u_j) v_i \, ds \right) \\ & - \int_{\Gamma} \frac{\partial u_1}{\partial n_1} v_1 \, ds - \int_{\Gamma} \frac{\partial u_2}{\partial n_2} v_2 \, ds = \sum_{i=1}^2 \int_{\Omega_i} f v_i \, dx, \end{aligned} \quad (16)$$

for all test functions (v_1, v_2) .

Comparing this with (12), shows the non conformity of the penalty method, i.e. the exact solution does not satisfy the penalty variational form. However, the above form would already be the rudimentary Nitsche.

The next step is to use the relation

$$\frac{\partial u_2}{\partial n_2} = -\frac{\partial u_1}{\partial n_1}, \quad (17)$$

to get the weak form

$$\begin{aligned} & \sum_{i=1}^2 \left(\int_{\Omega_i} \nabla u_i \cdot \nabla v_i \, dx + \int_{\Gamma} \frac{1}{\varepsilon} (u_i - u_j) v_i \, ds \right) \\ & - \int_{\Gamma} \frac{\partial u_1}{\partial n_1} (v_1 - v_2) \, ds = \sum_{i=1}^2 \int_{\Omega_i} f v_i \, dx. \end{aligned} \quad (18)$$

However, this is not entirely satisfying, since we end up with a non symmetric method for a problem that is symmetric.

Hence, the final step is to symmetrise the problem. For the exact solution $u = (u_1, u_2)$ it holds $u_1 = u_2$ on Γ , and hence it holds

$$\int_{\Gamma} \frac{\partial v_1}{\partial n_1} (u_1 - u_2) \, ds = 0. \quad (19)$$

We can thus add this to the variational form, and we conclude that the exact solution $u = (u_1, u_2)$ satisfies

$$\begin{aligned} & \sum_{i=1}^2 \left(\int_{\Omega_i} \nabla u_i \cdot \nabla v_i \, dx + \int_{\Gamma} \frac{1}{\varepsilon} (u_i - u_j) v_i \, ds \right) \\ & - \int_{\Gamma} \frac{\partial u_1}{\partial n_1} (v_1 - v_2) \, ds - \int_{\Gamma} \frac{\partial v_1}{\partial n_1} (u_1 - u_2) \, ds = \sum_{i=1}^2 \int_{\Omega_i} f v_i \, dx \end{aligned} \quad (20)$$

for all test functions $v = (v_1, v_2)$.

By denoting

$$\begin{aligned} \mathcal{B}(u, v) &= \sum_{i=1}^2 \left(\int_{\Omega_i} \nabla u_i \cdot \nabla v_i \, dx + \int_{\Gamma} \frac{1}{\varepsilon} (u_i - u_j) v_i \, ds \right) \\ & - \int_{\Gamma} \frac{\partial u_1}{\partial n_1} (v_1 - v_2) \, ds - \int_{\Gamma} \frac{\partial v_1}{\partial n_1} (u_1 - u_2) \, ds, \end{aligned} \quad (21)$$

we define the *Nitsche's method*: find $u_h = (u_{1,h}, u_{2,h}) \in V_h$ such that

$$\mathcal{B}(u_h, v) = \sum_{i=1}^2 \int_{\Omega_i} f v_i \, dx \quad \forall v \in V_h. \quad (22)$$

This is the one-sided Nitsche, a master slave method, with the domain Ω_1 being the slave.

By construction, the method is *consistent*. It remains to choose ε so that the method is stable and have an optimal order of convergence. To this end we first have

$$\mathcal{B}(v, v) = \sum_{i=1}^2 \int_{\Omega_i} |\nabla v_i|^2 \, dx - 2 \int_{\Gamma} \frac{\partial v_1}{\partial n_1} (v_1 - v_2) \, ds + \int_{\Gamma} \frac{1}{\varepsilon} (v_1 - v_2)^2 \, ds. \quad (23)$$

Schwarz inequality gives

$$\begin{aligned} \left| \int_{\Gamma} \frac{\partial v_1}{\partial n_1} (v_1 - v_2) ds \right| &= \left| \int_{\Gamma} \varepsilon^{1/2} \frac{\partial v_1}{\partial n_1} \varepsilon^{-1/2} (v_1 - v_2) ds \right| \\ &\leq \left(\int_{\Gamma} \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds \right)^{1/2} \left(\int_{\Gamma} \frac{1}{\varepsilon} (v_1 - v_2)^2 ds \right)^{1/2}. \end{aligned} \quad (24)$$

Using the inequality $ab \leq a^2 + \frac{1}{4}b^2$, $a, b \in \mathbb{R}$, we estimate

$$\left(\int_{\Gamma} \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds \right)^{1/2} \left(\int_{\Gamma} \frac{1}{\varepsilon} (v_1 - v_2)^2 ds \right)^{1/2} \leq \int_{\Gamma} \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds + \frac{1}{4} \int_{\Gamma} \frac{1}{\varepsilon} (v_1 - v_2)^2 ds. \quad (25)$$

Combining (24) and (25), then gives

$$-2 \int_{\Gamma} \frac{\partial v_1}{\partial n_1} (v_1 - v_2) ds \geq -2 \int_{\Gamma} \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds - \int_{\Gamma} \frac{1}{2\varepsilon} (v_1 - v_2)^2 ds, \quad (26)$$

which together with (23) gives

$$\mathcal{B}(v, v) \geq \sum_{i=1}^2 \int_{\Omega_i} |\nabla v_i|^2 dx - 2 \int_{\Gamma} \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds + \int_{\Gamma} \frac{1}{2\varepsilon} (v_1 - v_2)^2 ds. \quad (27)$$

Now, let $K \subset \Omega_1$ be an element with a side/face $E \subset \Gamma$. By a scaling argument there exists a positive constant C_I such that

$$C_I \int_K |\nabla v_1|^2 dx \geq h_E \int_E \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds, \quad (28)$$

where h_E is the local mesh length at E . This gives

$$\int_K |\nabla v_1|^2 dx - 2 \int_E \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds \geq \left(1 - 2\varepsilon \frac{C_I}{h_E} \right) \int_K |\nabla v_1|^2 dx. \quad (29)$$

Hence, if we choose ε such that $\varepsilon|_E = h_E/2\gamma$, with $\gamma > C_I$, it holds

$$\int_K |\nabla v_1|^2 dx - 2 \int_E \varepsilon \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds \gtrsim \int_K |\nabla v_1|^2 dx. \quad (30)$$

Combining with (23), then yields the result. There exists a constant C_I , so that the choice $\varepsilon|_E = h_E/2\gamma$, with $\gamma > C_I$, for all edges/faces on Γ , yields the *stability estimate*

$$\mathcal{B}(v, v) \gtrsim \sum_{i=1}^2 \int_{\Omega_i} |\nabla v_i|^2 dx + \sum_{E \subset \Gamma} \int_{\Gamma} \frac{1}{h_E} (v_1 - v_2)^2 ds = \|v\|_E^2 \quad (31)$$

By construction, the method is consistent, and together with the stability, one obtains an optimal *a priori error estimate*:

$$\|u - u_h\|_E \lesssim \inf_{v \in V_h} \|u - v\|_E \lesssim h^{s-1} \|u\|_s, \quad 1 \leq s \leq p+1. \quad (32)$$

Remark. When implementing the method, the crucial question is of course how difficult it is to estimate the constant C_I . For piecewise linear elements this is simple. On the element K , with edge/face E , ∇v_1 is a constant vector. Hence, it holds

$$\int_K |\nabla v_1|^2 dx = |\nabla v_1|^2 \int_K dx \quad \text{and} \quad \int_E \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds = \left| \frac{\partial v_1}{\partial n_1} \right|^2 \int_E ds. \quad (33)$$

Since

$$|\nabla v_1|^2 \geq \left| \frac{\partial v_1}{\partial n_1} \right|^2, \quad (34)$$

it holds

$$\int_K |\nabla v_1|^2 dx \geq \frac{\int_K dx}{\int_E ds} \int_E \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds. \quad (35)$$

Since $\int_K dx$ is the volume/area of K , and $\int_E ds$ is the area/length of E , respectively, the ratio $\int_K dx / \int_E ds$ is proportional to the diameter h_E , as the elements are assumed to be shape regular.

For linear elements the threshold value for the stability is then $\int_K dx / 2 \int_E ds$ and we thus have uniform stability by choosing

$$\varepsilon|_E = \alpha \frac{\int_K dx}{2 \int_E ds}, \quad (36)$$

with a chosen constant $\alpha < 1$.

In [17] the following explicit estimate for the inequality (28) is given

$$\frac{n \int_K dx}{(p+1)(p+n) \int_E ds} \int_K |\nabla v_1|^2 dx \geq \int_E \left| \frac{\partial v_1}{\partial n_1} \right|^2 ds, \quad (37)$$

where p is the polynomial degree and n is the space dimension. $\square$

In the *a posteriori estimator* the interface gives the terms

$$h_E \left\| \frac{\partial u_{1,h}}{\partial n_1} + \frac{\partial u_{2,h}}{\partial n_2} \right\|_{0,E}^2 \quad (38)$$

and

$$h_E \|u_{1,h} - u_{2,h}\|_{0,E}^2, \quad (39)$$

for an edge/face $E \subset \Gamma$. The estimator can be shown to be both reliable and efficient [6].

In the formulation (21)–(22), the role of u_1 and u_2 can be changed. The most common is, however, to use the average

$$\begin{aligned} \mathcal{B}(u, v) = & \sum_{i=1}^2 \left(\int_{\Omega_i} \nabla u_i \cdot \nabla v_i dx + \int_{\Gamma} \frac{1}{\varepsilon} (u_i - u_j) v_i ds \right) \\ & - \int_{\Gamma} \frac{1}{2} \left(\frac{\partial u_1}{\partial n_1} + \frac{\partial u_2}{\partial n_2} \right) (v_1 - v_2) ds \\ & - \int_{\Gamma} \frac{1}{2} \left(\frac{\partial v_1}{\partial n_1} + \frac{\partial v_2}{\partial n_2} \right) (u_1 - u_2) ds. \end{aligned} \quad (40)$$

For the case when the material parameters are different in the two domains, a convex combination of fluxes can be used, cf. [6, 7].

Numerical verifications

In this section we will numerically confirm the theoretical results. Our first test problem has the smooth solution

$$u(x, y) = xy \sin \frac{\pi x}{2} \sin \pi y \quad (41)$$

in the domains

$$\Omega_1 = (0, 1) \times (0, 1) \quad \text{and} \quad \Omega_2 = (1, 2) \times (0, 1). \quad (42)$$

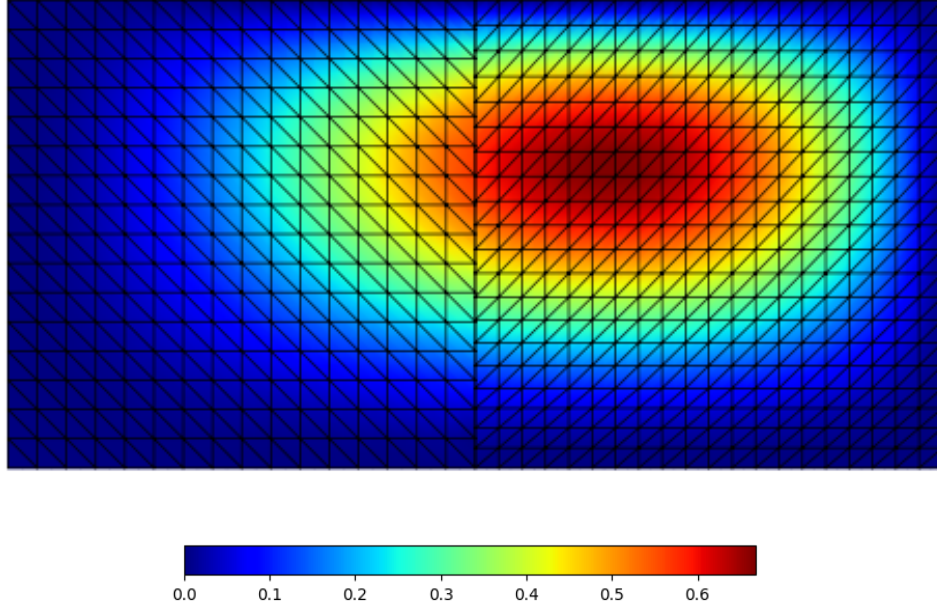


Figure 1. The smooth test problem.

We use uniform refinements with initial meshes that do not match at the interface and linear element, $p = 1$. In the penalty method we use $\epsilon \approx h$. Figure 2 shows the convergence with respect to the mesh length h . As predicted, the rates are $\mathcal{O}(h^{1/2})$ and $\mathcal{O}(h)$ for the penalty method and Nitsche's method, respectively.

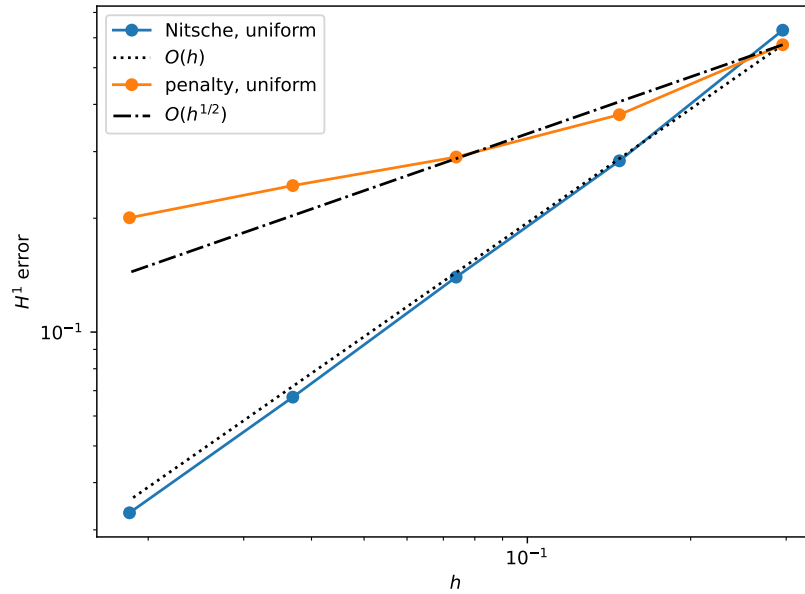


Figure 2. The convergence rates for the smooth problem and a uniform mesh refinement.

Next, we run the adaptive algorithms. We see from Figure 3, that the adaptive method detects the low convergence rate at the interface for the penalty method.

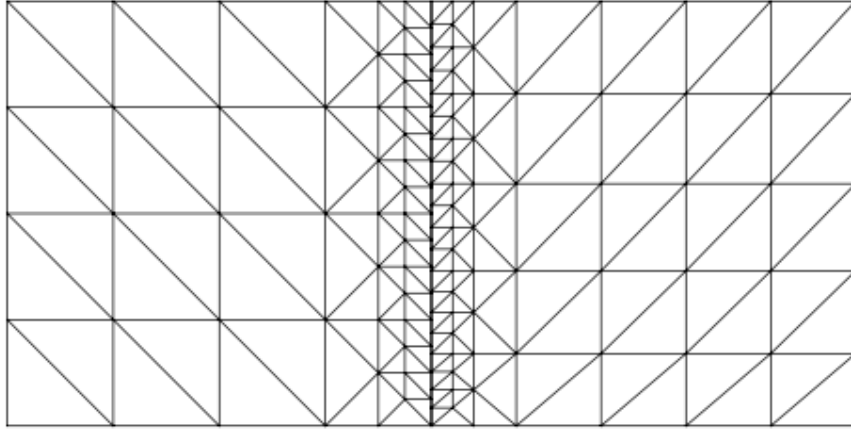


Figure 3. The adaptive mesh for the penalty method.

For Nitsche's method, any extra refinement along the interface does not occur, see Figure 4.

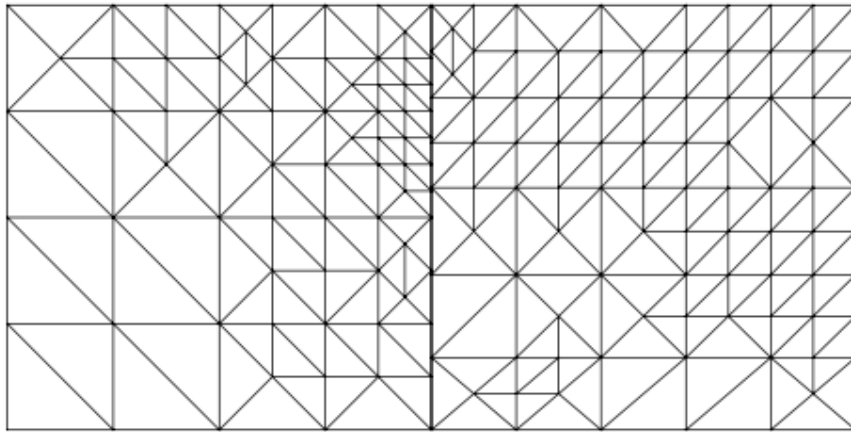


Figure 4. The adaptive mesh for the Nitsche's method.

Figure 5 shows that even with adaptivity the penalty falls short compared to Nitsche's method. Now the convergence rates are with respect to the number of degrees of freedom N .

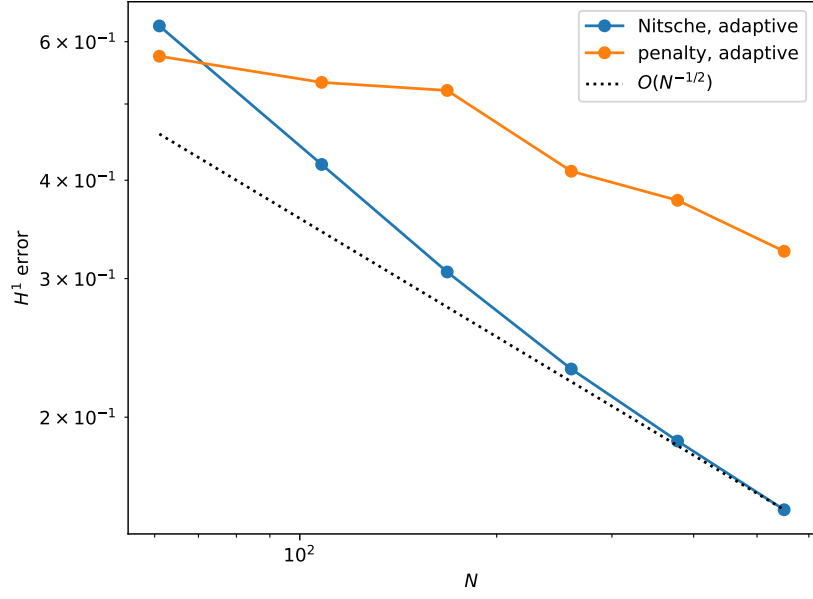


Figure 5. The convergence rates for the adaptive algorithms.

The second test example is a problem with a singularity. The domain is L-shaped and the solution is $r^{2/3}$. The solution is in $H^{5/2}$ and we choose second degree polynomials, i.e., $p = 2$ in order to enforce the adaptivity. The adaptive Nitsche's method detects the corner singularity and refines accordingly.

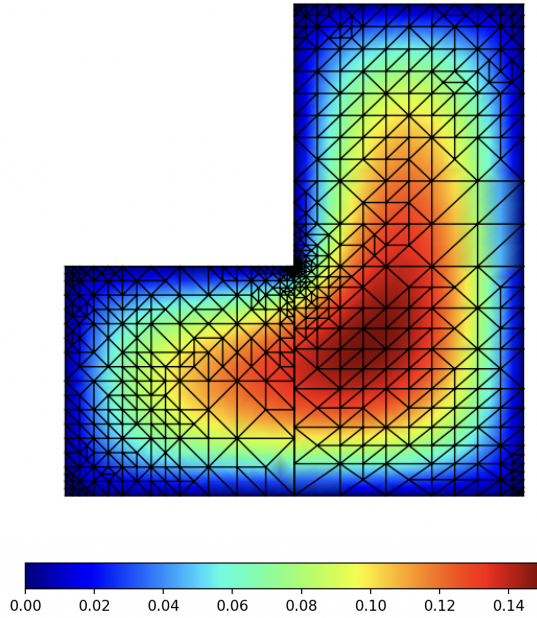


Figure 6. The adaptive Nitsche's method for the singular test problem.

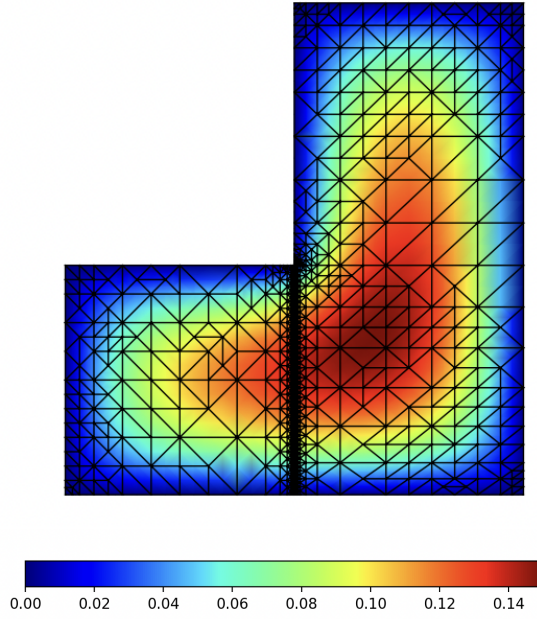


Figure 7. The adaptive penalty method for the singular test problem.

Here again, we see that the modelling error in the penalty method leads to a refinement along the whole interface.

In the last figure the convergence rates are plotted and we again see the superiority of Nitsche's method.

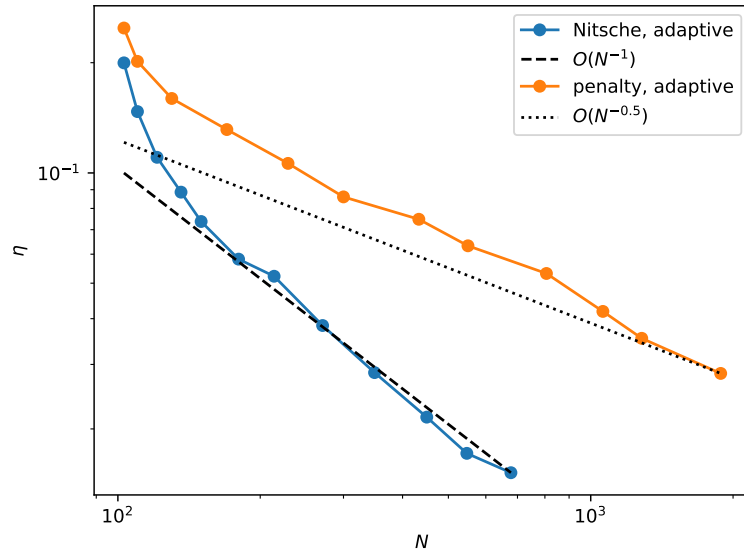


Figure 8. The convergence rates for the singular test problem.

Concluding remarks

We have shown that the Nitsche's method can be seen as a very simply modification of the penalty method, yielding a method with optimal order of convergence. One can say that

it is pity that Nitsche published his paper in German and in a not well-spread journal. As a consequence, it took a long time for the method to spread.

However, once one understands the method, one is stunned by its simplicity and elegance, and it is clear that there is an abundance of applications.

Let us close the paper by some recollections by the senior author of this paper. He stumbled on the method by studying stabilized mixed methods for enforcing boundary conditions, and concluded that these are essentially the same as Nitsche. At the Finnish Mechanics Days 1994, in Jyväskylä he presented a paper [13]. The paper was later expanded and published internationally [14], and it ended with the conclusion that the method should be explored for contact problems. This was done by Jouni Freund in his thesis [4] and in a short note in the world congress of computational mechanics in Buenos Aires 1998 [15]. Peter Hansbo was the opponent at Freund's thesis defence, and later Hansbo and his Swedish colleagues, Mats Larson and Erik Burman, have studied the method intensively for a variety of problems.

Another group that has made substantial contributions for elastic contact consists of Yves Renard, Patrick Hild and Franz Chouly, summarized in the recent monograph [2]. This book also include, e.g., frictional and dynamic problems.

Nitsche's method is by now included in several FEA codes, like COMSOL (<https://www.comsol.com>) and GetFEM (<https://getfem.org>).

References

- [1] Ivo Babuška. The finite element method with penalty. *Math. Comp.*, 27:221–228, 1973. <https://doi.org/10.2307/2005611>
- [2] Franz Chouly, Patrick Hild, and Yves Renard. *Finite element approximation of contact and friction in elasticity*, volume 48 of *Advances in Mechanics and Mathematics*. Birkhäuser/Springer, Cham, [2023] ©2023. Advances in Continuum Mechanics. <https://doi.org/10.1007/978-3-031-31423-0>
- [3] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.*, 49:1–23, 1943. <https://doi.org/10.1090/S0002-9904-1943-07818-4>
- [4] Jouni Freund. Space-time finite element methods for second order problems: an algorithmic approach. *Acta Polytech. Scand. Math. Comput. Manage. Eng. Ser.*, 79, 1996.
- [5] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical optimization*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1981.
- [6] Tom Gustafsson, Rolf Stenberg, and Juha Videman. Error analysis of Nitsche's mortar method. *Numer. Math.*, 142(4):973–994, 2019. <https://doi.org/10.1007/s00211-019-01039-5>
- [7] Tom Gustafsson, Rolf Stenberg, and Juha Videman. On Nitsche's method for elastic contact problems. *SIAM J. Sci. Comput.*, 42(2):B425–B446, 2020. <https://doi.org/10.1137/19M1246869>
- [8] Mika Juntunen and Rolf Stenberg. Nitsche's method for general boundary conditions. *Math. Comp.*, 78(267):1353–1374, 2009. <https://doi.org/10.1090/S0025-5718-08-02183-2>

- [9] M. Křížek, P. Neittaanmäki, and R. Stenberg, editors. *Finite Element Methods. Fifty years from the Courant element*, volume 164 of *Lecture notes in pure and applied mathematics*. Marcel Dekker, 1994.
- [10] Tod A. Laursen. *Computational contact and impact mechanics*. Springer-Verlag, Berlin, 2002. Fundamentals of modeling interfacial phenomena in nonlinear finite element analysis.
- [11] J. Nitsche. über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Sem. Univ. Hamburg*, 36:9–15, 1971. <https://doi.org/10.1007/BF02995904>
- [12] E. Polak. *Computational methods in optimization. A unified approach*, volume Vol. 77 of *Mathematics in Science and Engineering*. Academic Press, New York-London, 1971.
- [13] Rolf Stenberg. New and old finite element methods with approximation of boundary conditions. In Raino A.E. Mäkinen and Pekka Neittaanmäki, editors, *Proceedings of the 5th Finnish Mechanics Days*, pages 191–198, 1994.
- [14] Rolf Stenberg. On some techniques for approximating boundary conditions in the finite element method. volume 63, pages 139–148. 1995. International Symposium on Mathematical Modelling and Computational Methods Modelling 94 (Prague, 1994). [https://doi.org/10.1016/0377-0427\(95\)00057-7](https://doi.org/10.1016/0377-0427(95)00057-7)
- [15] Rolf Stenberg. Mortaring by a method of J.A. Nitsche. In E. Onate S Idelsohn and E. Dvorkin, editors, *COMPUTATIONAL MECHANICS, New Trends and Applications*. @CIMNE, Barcelona, Spain, 1998. <http://math.aalto.fi/~rstenber/Publications/wccm-mortar.pdf>
- [16] Gilbert Strang and George J. Fix. *An analysis of the finite element method*. Prentice-Hall Series in Automatic Computation. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.
- [17] T. Warburton and J. S. Hesthaven. On the constants in *hp*-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Engrg.*, 192(25):2765–2773, 2003. [https://doi.org/10.1016/S0045-7825\(03\)00294-9](https://doi.org/10.1016/S0045-7825(03)00294-9)
- [18] Peter Wriggers. *Computational Contact Mechanics*. Springer, 2 edition, 2006.

Rolf Stenberg
 Department of Mathematics and Systems Analysis
 Aalto University

Tom Gustafsson
 Department of Energy and Mechanical Engineering
 Aalto University