Chapter 1

Nitsche's Method: Step-by-Step Construction

Remark 1.1. Nitsche's method constitutes a well-established approach for weak enforcement of boundary conditions in partial differential equations (PDEs), which possess several advantageous theoretical properties. These include preservation of variational consistency and yielding symmetric positive-definite discrete linear systems. Consequently, it forms the foundation for geometrically unfitted finite element methods such as cut-FEM and extended FEM.

This presentation mainly refers to the comprehensive review of the literature by Benzaken et al. [1] regarding recent developments in Nitsche's method.

1.1 Variational Formulation and Galerkin Method

Remark 1.2. The discussion of Nitsche's method begins with the Poisson equation, followed by a quick review on constructing the corresponding variational problem and applying the Galerkin method for discretization.

Definition 1.1. The *model problem* for our exposition of Nitsche's method is the *strong form* of the scalar Poisson problem,

$$(S^P)\begin{cases} \text{Find } u : \overline{\Omega} \to \mathbb{R} \text{ such that} \\ \qquad \begin{aligned} -\Delta u &= \mathrm{f} \quad \text{in } \Omega \\ u &= \mathrm{g} \quad \text{on } \Gamma. \end{aligned} \end{cases}$$

with $\Omega \subset \mathbb{R}^d (d = 2, 3)$ a bounded domain with Lipschitz-continuous boundary $\Gamma = \partial\Omega$, $\mathrm{f} \in L^2(\Omega)$, and $\mathrm{g} \in H^{1/2}(\Gamma)$. The P superscript is used for the Poisson problem.

Lemma 1.2. Denote

$$\mathcal{V}^P \equiv H^1(\Omega), \tag{1.1}$$

$$\mathcal{V}^P_{\mathrm{g}} \equiv \left\{ v \in \mathcal{V}^P : v|_\Gamma = \mathrm{g} \right\}. \tag{1.2}$$

For $u \in \mathcal{V}^P_{\mathrm{g}}$ and $v \in \mathcal{V}^P_0$, we have

$$-\int_\Omega v \Delta u \; d\Omega = \int_\Omega \nabla u \cdot \nabla v \; d\Omega - \int_\Gamma (\nabla u \cdot \mathbf{n})\, v \; d\Gamma, \tag{1.3}$$

where $\mathbf{n}$ is the outward-facing unit normal to Ω.

Definition 1.3. The *variational formulation* of the model problem (S^P) writes,

$$(V^P)\begin{cases} \text{Find } u \in \mathcal{V}^P_{\mathrm{g}} \text{ such that} \\ \qquad a^P(u, v) = \langle f^P, v \rangle \\ \text{for every } v \in \mathcal{V}^P_0. \end{cases}$$

[1] Benzaken, J., Evans, J. A., & Tamstorf, R. (2024). Constructing Nitsche's method for variational problems. Archives of Computational Methods in Engineering

with the bilinear form $a^P(w,v)$ given by

$$a^P(w,v) \equiv \int_\Omega \nabla w \cdot \nabla v \; d\Omega \tag{1.4}$$

for $w \in \mathcal{V}_\mathrm{g}^P$ and $v \in \mathcal{V}_0^P$, and the linear functional f^P given by

$$\langle f^P, v\rangle \equiv \int_\Omega \mathrm{f} v \; d\Omega \tag{1.5}$$

for $v \in \mathcal{V}_0^P$.

Remark 1.3. The variational problem (V^P) also corresponds to the first-order optimality conditions of the *minimization problem*,

$$(M^P)\begin{cases} \text{Find } u \in \mathcal{V}_\mathrm{g}^P \text{ minimizing the total energy} \\ E_{\text{total}}^P(u) = \dfrac{1}{2}\displaystyle\int_\Omega \nabla u \cdot \nabla u \; d\Omega - \int_\Omega \mathrm{f} u \; d\Omega. \end{cases}$$

In fact, the term "variational form" originates from setting the first variation of the energy in the problem (M^P) to zero, from which the problem (V^P) is derived.

Suppose Ω can be represented by a *mesh* $\mathcal{K}$ of non-overlapping polygons (elements), such that

$$\Omega = \mathrm{int}\left(\overline{\bigcup_{K\in\mathcal{K}} K}\right).$$

For each element $K \in \mathcal{K}$, define the element size $h_K = \mathrm{diam}(K)$ and the mesh size $h = \max_{K\in\mathcal{K}} h_K$.

For a mesh $\mathcal{K}$, denote the *discrete space* $\mathcal{V}_h^P \subset \mathcal{V}^P$ consists of at least C^0-continuous piecewise polynomial approximations over the mesh $\mathcal{K}$.

The *discrete trial space* $\mathcal{V}_{\mathrm{g},h}^P$ is defined as the subspace of $\mathcal{V}_h^P$ that satisfies the Dirichlet boundary conditions on the edge mesh,

$$\mathcal{V}_{\mathrm{g},h}^P = \mathcal{V}_\mathrm{g}^P \cap \mathcal{V}_h^P.$$

The *discrete test space* $\mathcal{V}_{0,h}^P$ is defined as the homogeneous counterpart,

$$\mathcal{V}_{0,h}^P = \mathcal{V}_0^P \cap \mathcal{V}_h^P.$$

Definition 1.4. The *discrete variational problem* for (V^P) is given by

$$(V_h^P)\begin{cases} \text{Find } u_h \in \mathcal{V}_{\mathrm{g},h}^P \text{ such that} \\ \qquad a^P(u_h, v_h) = \langle f^P, v_h\rangle \\ \text{for every } v_h \in \mathcal{V}_{0,h}^P. \end{cases}$$

Therefore, we say (V_h^P) a *Galerkin Method.*

The Galerkin method exhibits three fundamental mathematical properties:

- **Symmetry**: $a^P(w_h, v_h) = a^P(v_h, w_h)$. This symmetry yields symmetric discretized linear systems, enabling efficient solvers like geometric multigrid.
- **Coercivity on $\mathcal{V}_{0,h}^P$**: There exists $C_1 > 0$ such that for all $w_h \in \mathcal{V}_{0,h}^P$,

  $$a^P(w_h, w_h) \ge C_1 \|w_h\|_{\mathcal{V}^P}.$$

  Coercivity guarantees unique solutions via the Lax-Milgram theorem.
- **Consistency**: For the exact solution u of (V^P),

  $$a^P(u, v_h) - \langle f^P, v_h\rangle = 0 \quad \forall v_h \in \mathcal{V}_{0,h}^P.$$

  Consistency ensures numerical accuracy via the Cea's lemma and induces Galerkin Orthogonality:

  $$\underbrace{0 = a^P(u - u_h, v_h)}_{\text{Galerkin Orthogonality}} = \underbrace{a^P(u, v_h) - a^P(u_h, v_h)}_{\text{Consistency}},$$

  demonstrating that the approximation error $u - u_h$ is a^P-orthogonal to $\mathcal{V}_{0,h}^P$.

Remark 1.4. Enforcing strong Dirichlet boundary conditions on discrete spaces $\mathcal{V}_{g,h}^P$ and $\mathcal{V}_{0,h}^P$ works easily for basic cases like simple approximation spaces or standard applications (e.g., the Poisson equation shown here). However, for advanced scenarios involving:

- Complex approximation methods (e.g., B-splines, subdivision surfaces)
- Challenging applications (e.g., fourth-order PDEs, moving boundary problems)

strict enforcement becomes much harder. This difficulty explains why weak enforcement of Dirichlet conditions is widely preferred for complex problems.

1.2 Coercivity of Penalty Method

Remark 1.5. The penalty method provides a simple way to weakly enforce boundary conditions. Instead of strictly enforcing boundary rules, it uses the full approximation space $\mathcal{V}_h^P$ for both test and trial functions. However, the bilinear form in (1.4) loses coercivity over $\mathcal{V}_h^P$ compared to $\mathcal{V}_{0,h}^P$. This failure to meet Lax-Milgram requirements stems from (1.4)'s non-trivial kernel over $\mathcal{V}_h^P$. To restore coercivity, residual-driven penalty terms are added to both sides of the variational equation.

We modify the finite-dimensional counterparts of (1.4) and (1.5) with a *penalty term*,

$$a_{\text{pen}}^P(w_h, v_h) \equiv \int_\Omega \nabla w_h \cdot \nabla v_h \, d\Omega + \frac{C_{\text{pen}}}{h} \int_\Gamma w_h v_h \, d\Gamma, \tag{1.6}$$

$$\left\langle f_{\text{pen}}^P, v_h \right\rangle \equiv \int_\Omega \mathrm{f} v_h \, d\Omega + \frac{C_{\text{pen}}}{h} \int_\Gamma \mathrm{g} v_h \, d\Gamma. \tag{1.7}$$

The constant $C_{\text{pen}} > 0$ is known as the *penalty parameter.* Then the *discrete variational penalty problem* is given by:

$$(V_{\text{pen}}^P) \begin{cases} \text{Find } u_h \in \mathcal{V}_h^P \text{ such that} \\ \qquad a_{\text{pen}}^P(u_h, v_h) = \left\langle f_{\text{pen}}^P, v_h \right\rangle \\ \text{for every } v_h \in \mathcal{V}_h^P. \end{cases}$$

Theorem 1.5. When $C_{\text{pen}} > 0$, the modified form a_{pen}^P from (1.6):

- Restores coercivity over the full space $\mathcal{V}_h^P$
- Preserves boundedness properties

Proof. As for boundedness, we have

$$\begin{aligned} a_{\text{Pen}}^P(w_h, w_h) &= (\nabla w_h, \nabla v_h)_\Omega + \frac{C_{\text{pen}}}{h} (w_h, v_h)_\Gamma \\ &\leq \|\nabla w_h\| \, \|\nabla v_h\| + \frac{C_{\text{pen}}}{h} \left(h^{-\frac{1}{2}} \|w_h\| + h^{\frac{1}{2}} \|\nabla w_h\| \right) \left(h^{-\frac{1}{2}} \|v_h\| + h^{\frac{1}{2}} \|\nabla v_h\| \right) \\ &\lesssim \|w_h\|_{H^1} \|v_h\|_{H^1}, \end{aligned}$$

where the second step follows from the discrete trace inequality.

As for coercivity, we prove it by contradiction. If not, then there exists a sequence $(w_h) \subset \mathcal{V}_h^P$, such that

$$\begin{cases} \|w_h^n\| = 1 \\ a_{\text{pen}}^P(w_h^n, w_h^n) \to 0 \end{cases}$$

The second relation leads to

$$\|\nabla w_h^n\| \to 0 \quad \text{and} \quad \|w_h^n\|_{L^2(\Gamma)} \to 0.$$

Since (w_h) is bounded in H^1-sense, and H^1 is compactly embedded into L^2, there exists the subsequence of (w_h), i.e. $(w_h)_{n_j}$, converges to some v in L^2-sense. Notice that both $(w_h)_{n_j}$ and $(\nabla w_h)_{n_j}$ are Cauchy sequence in L^2-sense, which yields $(w_h)_{n_j}$ is Cauchy in H^1-sense, and $(w_h)_{n_j}$ converges to some $w \in H^1$ due to the completeness of H^1.

Therefore,

$$\|\nabla w\| = \lim_{j \to \infty} \|\nabla {w_h}^{n_j}\| = 0,$$

which means w is constant on Ω (a.e.). Furthermore, $w \equiv 0$ due to $\|\mathrm{Tr}(w)\|_{L^2(\Gamma)} = 0$. However, this contradicts $\|w_h^{n_j}\| = 1$. ∎

The penalty method is essentially introducing an additional "energy" term that enforces the boundary condition. The constant C_{pen}, known as the penalty parameter, is typically chosen heuristically or experimentally based on the importance of enforcing the Dirichlet boundary conditions. The explicit h-dependence in the penalty parameter comes from a dimensional argument, which will be made later in this section.

Remark 1.6. As $C_{\text{pen}} \to \infty$, the penalty method asymptotically recovers the strong satisfaction of the Dirichlet boundary conditions. While a large penalty parameter ensures stronger enforcement of boundary conditions, it can lead to ill-conditioning of the resulting matrix system after discretization, thereby adversely affecting numerical performance. Hence, **the choice of C_{pen} involves a trade-off between boundary condition enforcement and numerical stability**.

While coercivity is preserved, the current formulation introduces inconsistency with the original problem.

Proposition 1.6. The form a^P_{pen} defined in (1.6) is not consistent.

Proof. To see this, suppose that u is smooth enough so that (1.3) holds and observe that the bilinear form is inconsistent since

$$\begin{aligned} a^P_{\text{pen}}(u, v_h) - a^P_{\text{pen}}(u_h, v_h) = a^P_{\text{pen}}(u, v_h) - \left\langle f^P_{\text{pen}}, v_h \right\rangle &= \int_\Omega \nabla u \cdot \nabla v_h \; d\Omega + \frac{C_{\text{pen}}}{h} \int_\Gamma u v_h \; d\Gamma \\ &- \int_\Omega \mathrm{f} v_h \; d\Omega - \frac{C_{\text{pen}}}{h} \int_\Gamma \mathrm{g} v_h \; d\Gamma \\ &= \int_\Gamma (\nabla u \cdot \mathbf{n})\, v_h \; d\Gamma + \frac{C_{\text{pen}}}{h} \int_\Gamma (u - \mathrm{g})\, v_h \; d\Gamma \\ &- \int_\Omega (\mathrm{f} + \Delta u)\, v_h \; d\Omega \\ &= \int_\Gamma (\nabla u \cdot \mathbf{n})\, v_h \; d\Gamma \neq 0, \end{aligned} \tag{1.8}$$

where the second equality follows from (1.6) and (1.7), and the residual integrals going from the third equality to the fourth vanish because $u_h = \mathrm{g}$ on Γ and $\mathrm{f} = -\Delta u_h$ in Ω. □

1.3 Consistency and Symmetry of Nitsche's Method

This inconsistency leads to an inaccurate numerical approximation. To restore variational consistency, we should address its root cause, i.e. the residual boundary term in (1.8). By incorporating this boundary integral (with opposite sign) into the penalty form (1.6), we recover a consistent formulation, and that is the foundation of Nitsche's method.

Definition 1.7. The consistent bilinear form is given for $w_h, v_h \in \mathcal{V}^P_h$, by adding a boundary term that compromises the term of the boundary integral in (1.3), referred to as the *consistency term*. That is,

$$a^P_{\text{con}}(w_h, v_h) \equiv \int_\Omega \nabla w_h \cdot \nabla v_h \; d\Omega \underbrace{- \int_\Gamma (\nabla w_h \cdot \mathbf{n})\, v_h \; d\Gamma}_{\text{Consistency Term}} \underbrace{+ \frac{C_{\text{pen}}}{h} \int_\Gamma w_h v_h \; d\Gamma}_{\text{Penalty Term}}. \tag{1.9}$$

Proposition 1.8. The bilinear form a^P_{con} defined in (1.9) is consistent.

One can show that this bilinear form is consistent with the original boundary conditions by the same steps as in (1.8). Later in Section 1.4, we will demonstrate that this new bilinear form is also coercive over $\mathcal{V}^P_h$, provided that C_{pen} is sufficiently large. However, we have lost the symmetry that was present in the initial Galerkin formulation (1.4) and the penalty formulation (1.6).

A discretization based on (1.7) and (1.9) generally achieves optimal theoretical convergence rates in the energy norm. However, the lack of symmetry in this method might cause sub-optimal convergence rates when using weaker error measurements like lower-order Sobolev norms. Additionally, this approach produces linear systems that are not symmetric, which makes it more challenging to choose efficient equation solvers such as geometric multigrid methods.

Definition 1.9. A residual-based boundary integral known as the *symmetry term*, is added in the consistent form to restore the original symmetry of the bilinear form. This term is constructed from the symmetric counterpart to the consistency term. Incorporation into both the bilinear and linear forms yields,

$$a^P_h(w, v) \equiv \int_\Omega \nabla w \cdot \nabla v \; d\Omega \underbrace{- \int_\Gamma (\nabla w \cdot \mathbf{n})\, v \; d\Gamma}_{\text{Consistency Term}} \underbrace{- \int_\Gamma (\nabla v \cdot \mathbf{n})\, w \; d\Gamma}_{\text{Symmetry Term}} \underbrace{+ \frac{C_{\text{pen}}}{h} \int_\Gamma w v \; d\Gamma}_{\text{Penalty Term}} \tag{1.10}$$

and

$$\langle f_h^P, v \rangle \equiv \int_\Omega \mathrm{f}v \; d\Omega \underbrace{- \int_\Gamma (\nabla v \cdot \mathbf{n})\, \mathrm{g} \; d\Gamma}_{\text{Symmetry Term}} \underbrace{+ \frac{C_{\text{pen}}}{h} \int_\Gamma \mathrm{g}v \; d\Gamma}_{\text{Penalty Term}}. \tag{1.11}$$

Then the *Nitsche's method* for the model problem (S^P) writes,

$$(N_h^P) \begin{cases} \text{Find } u_h \in \mathcal{V}_h^P \text{ such that} \\ \qquad\qquad a_h^P(u,v) = \langle f_h^P, v \rangle \\ \text{for every } v \in \mathcal{V}_h^P. \end{cases}$$

The bilinear form $a_h^P(\cdot,\cdot)$ defined in (1.10) is inherently consistent and symmetric. The next subsection will prove that such bilinear form is coercive.

1.4 Ensuring Coercivity through Penalty Parameter

In Section 1.3, we have asserted that adding penalty terms restores the coercivity of the bilinear form in (1.10). This subsection outlines the selection of the penalty parameter, guided by several key inequalities.

Lemma 1.10. A required inequality is the *Cauchy-Schwarz inequality* on the boundary, given by:

$$\int_\Gamma |(\nabla w_h \cdot \mathbf{n})\, w_h| \, d\Gamma \leq \|\nabla w_h \cdot \mathbf{n}\|_{0,\Gamma} \|w_h\|_{0,\Gamma},$$

where

$$\|f\|_{0,\Gamma}^2 \equiv \int_\Gamma |f|^2 \, d\Gamma \tag{1.12}$$

is the L^2-norm on Γ.

The Cauchy-Schwarz inequality bounds a norm of a product by the product of norms. It will be used to bound duality pairings involving boundary quantities.

Lemma 1.11. Another required inequality is the *discrete trace inequality*, which takes the form:

$$\|\nabla w_h \cdot \mathbf{n}\|_{0,\Gamma}^2 \leq \frac{C_{\text{tr}}}{h} \|\nabla w_h\|_{0,\Omega}^2. \tag{1.13}$$

This establishes a proportional relationship between boundary values and interior values through a constant factor combined with mesh-dependent scaling.

The constant C_{tr} depends on the geometry of the domain Ω. It is convenient to express its dependence on the mesh size h explicitly, as shown in (1.13). For instance,

$$\|\nabla w_h \cdot \mathbf{n}\|_{0,\Gamma}^2 = \int_\Gamma \underbrace{|\nabla w_h \cdot \mathbf{n}|^2}_{(\mathcal{O}(h^{-1}))^2} \underbrace{d\Gamma}_{\mathcal{O}(h^{d-1})} \leq C \int_\Omega \underbrace{|\nabla w_h|^2}_{(\mathcal{O}(h^{-1}))^2} \underbrace{d\Omega}_{\mathcal{O}(h^d)} = C \|\nabla w_h\|_{0,\Omega}^2.$$

Therefore, one can conclude from the mesh scaling argument that $\|\nabla w_h \cdot \mathbf{n}\|_{0,\Gamma}^2 = \mathcal{O}(h^{d-3})$ and $\|\nabla w_h\|_{0,\Omega}^2 = \mathcal{O}(h^{d-2})$, hence the coefficient C must contain an additional factor of h^{-1} and we write $C = C_{\text{pen}}/h$.

Theorem 1.12. The form a_h^P is coercive if a proper penalty parameter C_{pen} is set.

Proof. Let $\gamma > 0$ and observe that:

$$
\begin{aligned}
a_h^P(w_h, w_h) &= \int_\Omega \nabla w_h \cdot \nabla w_h \, d\Omega - 2\int_\Gamma (\nabla w_h \cdot \mathbf{n})\, w_h \, d\Gamma + \frac{C_{\text{pen}}}{h}\int_\Gamma w_h^2 \, d\Gamma \\
&= \|\nabla w_h\|_{0,\Omega}^2 - 2\int_\Gamma (\nabla w_h \cdot \mathbf{n})\, w_h \, d\Gamma + \frac{C_{\text{pen}}}{h}\|w_h\|_{0,\Gamma}^2 \\
&\geq \|\nabla w_h\|_{0,\Omega}^2 - 2\|\nabla w_h \cdot \mathbf{n}\|_{0,\Gamma}\|w_h\|_{0,\Gamma} + \frac{C_{\text{pen}}}{h}\|w_h\|_{0,\Gamma}^2 \\
&\geq \|\nabla w_h\|_{0,\Omega}^2 - 2\left(\frac{1}{2\gamma}\|\nabla w_h \cdot \mathbf{n}\|_{0,\Gamma}^2 + \frac{\gamma}{2}\|w_h\|_{0,\Gamma}^2\right) + \frac{C_{\text{pen}}}{h}\|w_h\|_{0,\Gamma}^2 \\
&\geq \|\nabla w_h\|_{0,\Omega}^2 - \left(\frac{C_{\text{tr}}}{h\gamma}\|\nabla w_h\|_{0,\Omega}^2 + \gamma\|w_h\|_{0,\Gamma}^2\right) + \frac{C_{\text{pen}}}{h}\|w_h\|_{0,\Gamma}^2 \\
&= \left(1 - \frac{C_{\text{tr}}}{h\gamma}\right)\|\nabla w_h\|_{0,\Omega}^2 + \left(\frac{C_{\text{pen}}}{h} - \gamma\right)\|w_h\|_{0,\Gamma}^2,
\end{aligned}
$$

where the third step follows from Lemma 1.10, the fourth step from Young's inequality, and the fifth step from Lemma 1.11.

Then the coercivity of the bilinear form $a_h^P(\cdot,\cdot)$ is ensured by choosing γ and C_{pen} such that,

$$
\gamma > \frac{C_{\text{tr}}}{h} \quad \text{and} \quad \frac{C_{\text{pen}}}{h} > \gamma.
$$

In other words, assume the penalty parameter C_{pen} is chosen to satisfy

$$
C_{\text{pen}} > C_{\text{tr}} > 0,
$$

then there always exists a proper γ such that

$$
a_h^P(w_h, w_h) \geq C^*\left(\|\nabla w_h\|_{0,\Omega}^2 + \|w_h\|_{0,\Gamma}^2\right),
$$

where $C^* > 0$ is independent of w_h. The rest of proof is just the same as the proof for coercivity in Theorem 1.5. □

Practically, choosing $\lambda = 2C_{\text{tr}}$ often works well if C_{tr} can be approximated. The penalty parameter increases proportionally as the mesh refines (scaling with $1/h$), similar to trace inequality behavior. This highlights that Nitsche's penalty parameters inherently depend on mesh resolution.

Remark 1.7. Nitsche's method can be interpreted as a refinement of the discrete penalty energy minimization problem (M_{pen}^P), where the formulation is augmented with an energy term enforcing variational consistency of boundary conditions. Formally,

$$
(M_{\text{nit}}^P)\left\{
\begin{aligned}
&\text{Given f, g, find } u_h \in \mathcal{V}_h^P \text{ that minimizes the total energy} \\
&\quad E_{\text{nit}}^P(u_h) = \frac{1}{2}\int_\Omega \nabla u_h \cdot \nabla u_h \, d\Omega - \int_\Omega \text{f} u_h \, d\Omega \\
&\qquad\qquad - \int_\Gamma (\nabla u_h \cdot \mathbf{n})\,(u_h - g)\, d\Gamma + \frac{C_{\text{pen}}}{2h}\int_\Gamma (u_h - g)^2 \, d\Gamma.
\end{aligned}
\right.
$$

1.5 Nitsche's Method for Abstract Variational Problems

In this section, we apply a more formal approach to the derivation of Nitsche's formulation.

Consider $\mathcal{V}$ and $\mathcal{Q}$ as Hilbert spaces equipped with inner products $(\cdot,\cdot)_\mathcal{V}$ and $(\cdot,\cdot)_\mathcal{Q}$, and corresponding norms $|\cdot|_\mathcal{V} = (\cdot,\cdot)_\mathcal{V}^{1/2}$ and $|\cdot|_\mathcal{Q} = (\cdot,\cdot)_\mathcal{Q}^{1/2}$. The dual spaces of $\mathcal{V}$ and $\mathcal{Q}$ are $\mathcal{V}^*$ and $\mathcal{Q}^*$, respectively, with duality pairings ${}_{\mathcal{V}^*}\langle\cdot,\cdot\rangle_\mathcal{V}$ and ${}_{\mathcal{Q}^*}\langle\cdot,\cdot\rangle_\mathcal{Q}$.

Define the trace operator $\mathcal{T}: \mathcal{V} \to \mathcal{Q}$ as a bounded, surjective linear map. For $g \in \mathcal{Q}$, define

$$
\mathcal{V}_g \equiv \{v \in \mathcal{V} : \mathcal{T}v = g\}.
$$

Let $a: \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ be a bounded, symmetric, positive semi-definite bilinear form satisfying the coercivity condition on the kernel of $\mathcal{T}$:

$$
a(v, v) \geq C|v|_\mathcal{V}^2 \quad \forall v \in \mathcal{V}_0,
$$

for some constant $C > 0$.

Then we introduce the following *abstract variational problem*:

$$
(V)\begin{cases}
\text{Given } f \in \mathcal{V}^* \text{ and } g \in \mathcal{Q}, \text{ find } u \in \mathcal{V}_g \text{ such that}\\[1ex]
\qquad\qquad a(u, \delta u) = {}_{\mathcal{V}^*}\langle f, \delta u\rangle_{\mathcal{V}}\\[1ex]
\text{for every } \delta u \in \mathcal{V}_0.
\end{cases}
$$

The Lax-Milgram theorem guarantees that Problem (V) has a unique solution $u \in \mathcal{V}$ that depends continuously on the input $f \in \mathcal{V}^*$ and $g \in \mathcal{Q}$.

Then let $\mathcal{V}_h \subset \mathcal{V}$ be a finite-dimensional approximation space, and define $\mathcal{V}_{g,h} = \mathcal{V}_h \cap \mathcal{V}_g$ for each $g \in \mathcal{Q}$. The Galerkin method of Problem (V) is given by:

$$
(V_h)\begin{cases}
\text{Given } f \in \mathcal{V}^* \text{ and } g \in \mathcal{Q}, \text{ find } u_h \in \mathcal{V}_{g,h}\\
\text{such that}\\[1ex]
\qquad a(u_h, \delta u_h) = {}_{\mathcal{V}^*}\langle f, \delta u_h\rangle_{\mathcal{V}}\\[1ex]
\text{for every } \delta u_h \in \mathcal{V}_{0,h}.
\end{cases}
\tag{1.13}
$$

According to the Lax-Milgram theorem, Problem (V_h) also has a unique solution $u_h \in \mathcal{V}_h$ that depends continuously on the input $f \in \mathcal{V}^*$ and $g \in \mathcal{Q}$. It is straightforward to show that the solution to Problem (V_h) best approximates the solution to Problem (V) with respect to the norm induced by the bilinear form $a(\cdot,\cdot)$.

Remark 1.8. The primary challenge with Problem (V_h) lies in the strong enforcement of the condition $\mathcal{T}u_h = g$. To address this, we employ Nitsche's method, as discussed in the previous section.

However, before introducing Nitsche's method for Problem (V), we must verify that two key assumptions are satisfied to ensure a formulation that is both provably stable and consistent. The first assumption involves the existence of a generalized Green's identity. The second assumption involves the existence of generalized trace and Cauchy-Schwarz inequalities.

Assumption 1.13. The following hold:

1. There exists a dense subspace $\tilde{\mathcal{V}} \subset \mathcal{V}$ with linear maps $\mathcal{L} : \tilde{\mathcal{V}} \to \mathcal{V}^*$ and $\mathcal{B} : \tilde{\mathcal{V}} \to \mathcal{Q}^*$ satisfying the *generalized Green's identity*:
$$a(w, v) = {}_{\mathcal{V}^*}\langle \mathcal{L}w, v\rangle_{\mathcal{V}} + {}_{\mathcal{Q}^*}\langle \mathcal{B}w, \mathcal{T}v\rangle_{\mathcal{Q}}, \quad \forall w \in \tilde{\mathcal{V}},\ v \in \mathcal{V}. \tag{1.14}$$

2. For $f \in \mathcal{V}^*$ and $g \in \mathcal{Q}$ with $u \in \tilde{\mathcal{V}}$, the solution u of Problem (V) satisfies $\mathcal{L}u = f$.

Such existence of a generalized Green's identity allows us to link the strong and weak formulations.

Assumption 1.14. There exist positive definite and self-adjoint operators $\eta : \mathrm{dom}(\eta) \subseteq \mathcal{Q}^* \to \mathcal{Q}$ and $\epsilon : \mathrm{dom}(\epsilon) \subseteq \mathcal{Q}^* \to \mathcal{Q}$, with ϵ surjective (hence invertible), such that:

1. The domain of definition of the operator $\mathcal{B} : \tilde{V} \to \mathcal{Q}^*$ can be extended to the enlarged space $\tilde{\mathcal{V}} + \mathcal{V}_h$ and the space $\{\mathcal{B}v \mid v \in \tilde{\mathcal{V}} + \mathcal{V}_h\}$ is a subset of $\mathrm{dom}(\eta)$.

2. *Generalized Trace Inequality*: For all $v_h \in \mathcal{V}_h$,
$${}_{\mathcal{Q}^*}\langle \mathcal{B}v_h, \eta\mathcal{B}v_h\rangle_{\mathcal{Q}} \le a(v_h, v_h).$$

3. *Generalized Cauchy-Schwarz Inequality*: For all $v, w \in \tilde{\mathcal{V}} + \mathcal{V}_h$,
$$\left|{}_{\mathcal{Q}^*}\langle \mathcal{B}v, \mathcal{T}w\rangle_{\mathcal{Q}}\right| \le \frac{1}{\gamma}\, {}_{\mathcal{Q}^*}\left\langle \mathcal{B}v, \eta\mathcal{B}v\right\rangle_{\mathcal{Q}}^{1/2} \cdot {}_{\mathcal{Q}^*}\left\langle \epsilon^{-1}\mathcal{T}w, \mathcal{T}w\right\rangle_{\mathcal{Q}}^{1/2}, \quad \gamma > 1.$$

These inequalities enable boundary norm control for Nitsche's method convergence analysis.

Remark 1.9. For the generalized Green's identity to hold, the solution to Problem (V) must be sufficiently smooth. This is why we introduced an additional subspace $\tilde{\mathcal{V}} \subset \mathcal{V}$ for which (1.14) holds.

Both η and ϵ can be interpreted as Riesz operators. The former is associated with a duality pairing related to the boundary operator $\mathcal{B}$, while the latter arises from a stabilized Lagrange multipliers method.

With the two assumptions above, we are ready to present Nitsche's method for the abstract variational problem given by Problem (V).

Definition 1.15 (Nitsche's method)**.** The *Nitsche's method* for the abstract variational problem (V) writes,

$$(N_h)\begin{cases}\text{Find } u_h \in \mathcal{V}_h \text{ such that} \\ \qquad\qquad a_h(u_h, \delta u_h) = \langle f_h, \delta u_h\rangle \\ \text{for every } \delta u_h \in \mathcal{V}_h.\end{cases}$$

with the bilinear form $a_h(\cdot,\cdot)$ given by

$$a_h(w,v) \equiv a(w,v) \underbrace{-_{\mathcal{Q}^*}\langle \mathcal{B}w, \mathcal{T}v\rangle_{\mathcal{Q}}}_{\text{Consistency Term}} \underbrace{-_{\mathcal{Q}^*}\langle \mathcal{B}v, \mathcal{T}w\rangle_{\mathcal{Q}}}_{\text{Symmetry Term}} \underbrace{+_{\mathcal{Q}^*}\langle \epsilon^{-1}\mathcal{T}v, \mathcal{T}w\rangle_{\mathcal{Q}}}_{\text{Penalty Term}}$$

for $w, v \in \tilde{\mathcal{V}} + \mathcal{V}_h$, and the linear functional $f_h(\cdot)$ given by

$$\langle f_h, \delta u_h\rangle \equiv {}_{\mathcal{V}^*}\langle f, \delta u_h\rangle_{\mathcal{V}} \underbrace{-_{\mathcal{Q}^*}\langle \mathcal{B}\delta u_h, g\rangle_{\mathcal{Q}}}_{\text{Symmetry Term}} \underbrace{+_{\mathcal{Q}^*}\langle \epsilon^{-1}\mathcal{T}\delta u_h, g\rangle_{\mathcal{Q}}}_{\text{Penalty Term}}$$

for $\delta u_h \in \mathcal{V}_h$.

Remark 1.10. Nitsche's method preserves several crucial properties that ensure its stability and convergence. To be specific, it is *consistent*, its bilinear form $a_h(\cdot,\cdot)$ is *symmetric*, and, given an appropriate choice of the parameter ϵ, the bilinear form $a_h(\cdot,\cdot)$ is also *coercive* on the discrete space $\mathcal{V}_h$.

Here, we present a fundamental result demonstrating the well-posedness of Nitsche's method along with Cea's error estimate.

Theorem 1.16 (Well-Posedness and Error Estimate)**.** Suppose Assumptions 1.13 and 1.14 hold. Then there exists a unique discrete solution $u_h \in \mathcal{V}_h$ to Problem (N_h).

Moreover, if the continuous solution $u \in \mathcal{V}$ to Problem (M) satisfies $u \in \tilde{\mathcal{V}}$, then the discrete solution u_h satisfies the following error estimate,

$$|||u - u_h||| \leq \left(1 + \frac{2}{1-\frac{1}{\gamma}}\right) \inf_{v_h \in \mathcal{V}_h} |||u - v_h|||,$$

where $|||\cdot||| : \tilde{\mathcal{V}} + \mathcal{V}_h \to \mathbb{R}$ is the energy norm defined by

$$|||v|||^2 \equiv a(v,v) + {}_{\mathcal{Q}^*}\langle \mathcal{B}v, \eta\mathcal{B}v\rangle_{\mathcal{Q}} + 2\,{}_{\mathcal{Q}^*}\langle \epsilon^{-1}\mathcal{T}v, \mathcal{T}v\rangle_{\mathcal{Q}}$$

and γ is given through Assumption 1.14.

Proof. First, we have

$$\begin{aligned} a_h(w,v) &\leq a(w,v) + |\langle \mathcal{B}w, \mathcal{T}v\rangle| + |\langle \mathcal{B}v, \mathcal{T}w\rangle| + \left\langle \epsilon^{-1}\mathcal{T}v, \mathcal{T}w\right\rangle \\ &\leq a(w,w)^{1/2}a(v,v)^{1/2} + \langle \mathcal{B}v, \eta\mathcal{B}v\rangle^{1/2}\langle \epsilon^{-1}\mathcal{T}w, \mathcal{T}w\rangle^{1/2} + \langle \mathcal{B}w, \eta\mathcal{B}w\rangle^{1/2}\langle \epsilon^{-1}\mathcal{T}v, \mathcal{T}v\rangle^{1/2} + \left\langle \epsilon^{-1}\mathcal{T}v, \mathcal{T}w\right\rangle \\ &\leq |||w|||\,|||v|||, \end{aligned}$$

for all $w, v \in \tilde{\mathcal{V}} + \mathcal{V}_h$, where the third step follows from applying a standard Cauchy-Schwarz inequality.

Also, similar to the proof in Theorem 1.12, applying the generalized Cauchy-Schwarz inequality of Assumption 1.14, Young's inequality, and the generalized trace inequality of Assumption 1.14 yields

$$a_h(v_h, v_h) \geq \left(1 - \frac{1}{\gamma}\right) a(u_h, v_h) + \left(1 - \frac{1}{\gamma}\right)\left\langle \epsilon^{-1}\mathcal{T}v_h, \mathcal{T}v_h\right\rangle \geq \frac{1}{2}\left(1 - \frac{1}{\gamma}\right)|||v_h|||^2.$$

Therefore, by the Lax-Milgram Theorem combined with the boundedness and coercivity of $a_h(\cdot,\cdot)$, Problem (N_h) is well-posed, i.e., there exists a unique discrete solution $u_h \in \mathcal{V}_h$ to Problem (N_h).

Next we give a Cea's error estimate. For $v_h \in \mathcal{V}_h$, since the Problem (V) is consistent with Problem (M), it follows that

$$|||u_h - v_h|||^2 \leq \frac{2}{1-\frac{1}{\gamma}} a_h(u_h - v_h, u_h - v_h) = \frac{2}{1-\frac{1}{\gamma}} a_h(u - v_h, u_h - v_h) \leq \frac{2}{1-\frac{1}{\gamma}} |||u - v_h|||\,|||u_h - v_h|||,$$

and thus

$$|||u_h - v_h||| \leq \frac{2}{1-\frac{1}{\gamma}} |||u - v_h|||.$$

Consequently,

$$|||u-u_h||| \leq |||u-v_h||| + |||u_h-v_h||| \leq \left(1+\frac{2}{1-\frac{1}{\gamma}}\right)|||u-v_h|||.$$

Since $v_h \in \mathcal{V}_h$ is arbitrary, the proof is completed. □

Remark 1.11. The above theorem applies to any formulation and problem setup for which Assumptions 1.13 and 1.14 hold. Consequently, constructing Nitsche's formulations for a new problem class should follow these steps:

Constructing Nitsche's Formulations

Step 1: Define an appropriate variational formulation, specifying the Hilbert spaces $\mathcal{V}$ and $\mathcal{Q}$, the mapping $\mathcal{T}:\mathcal{V}\to\mathcal{Q}$, and the bilinear form $a:\mathcal{V}\times\mathcal{V}\to\mathbb{R}$.

Step 2: Develop the generalized Green's identity

$$a(w,v) = {}_{\mathcal{V}^*}\langle \mathcal{L}w, v\rangle_{\mathcal{V}} + {}_{\mathcal{Q}^*}\langle \mathcal{B}w, \mathcal{T}v\rangle_{\mathcal{Q}}$$

by identifying the space $\tilde{\mathcal{V}}$ and the linear mappings $\mathcal{L}:\tilde{\mathcal{V}}\to\mathcal{V}^*$ and $\mathcal{B}:\tilde{\mathcal{V}}\to\mathcal{Q}^*$, in accordance with Assumption 1.13.

Step 3: Establish the generalized Cauchy-Schwarz inequality

$$|{}_{\mathcal{Q}^*}\langle \mathcal{B}v, \mathcal{T}w\rangle_{\mathcal{Q}}| \leq \frac{1}{\gamma}\left({}_{\mathcal{Q}^*}\langle \mathcal{B}v, \eta\mathcal{B}v\rangle_{\mathcal{Q}}^{\frac{1}{2}}\right)\left({}_{\mathcal{Q}^*}\langle \epsilon^{-1}\mathcal{T}w, \mathcal{T}w\rangle_{\mathcal{Q}}^{\frac{1}{2}}\right)$$

and the generalized trace inequality

$${}_{\mathcal{Q}^*}\langle \mathcal{B}v_h, \eta\mathcal{B}v_h\rangle_{\mathcal{Q}} \leq a(v_h, v_h)$$

by constructing proper linear mappings $\epsilon:\mathrm{dom}(\epsilon)\subseteq\mathcal{Q}^*\to\mathcal{Q}$ and $\eta:\mathrm{dom}(\eta)\subseteq\mathcal{Q}^*\to\mathcal{Q}$, in accordance with Assumption 1.14.

Step 4: Formulate Nitsche's method as described in Problem (N_h).

Example 1.17. We revisit Nitsche's formulations for the Poisson problem (S^P) introduced in the very beginning. However, this time we start from the abstract framework. To be specific:

- Lemma 1.2 corresponds to the generalized Green's identity, with

$$\begin{aligned}\mathcal{L}(w) &\equiv \text{Negative Laplacian Operator},\\ \mathcal{B}(w) &\equiv \text{Boundary Normal Trace Operator},\\ \mathcal{T}(v) &\equiv \text{Boundary Trace Operator}.\end{aligned}$$

- Lemma 1.11 corresponds to the generalized trace inequality with $\eta \equiv h/C_{\mathrm{tr}}$.
- Lemma 1.10 corresponds to the generalized Cauchy-Schwarz inequality with $\epsilon^{-1}\equiv C_{\mathrm{pen}}/h$ and $\gamma^2 = C_{\mathrm{pen}}/C_{\mathrm{tr}}$.

Therefore, Theorem 1.16 can be applied to Problem (N_h^P).

Corollary 1.18 (Well-Posedness and Error Estimate for the Poisson Problem)**.** Let $C_{\mathrm{pen}} = \gamma^2 C_{\mathrm{tr}}$, where $\gamma > 1$.

Then there exists a unique discrete solution $u_h \in \mathcal{V}_h^P$ to Problem (N_h^P). Moreover, if the continuous solution $u\in\mathcal{V}^P$ to Problem (M^P) satisfies $u\in\tilde{\mathcal{V}}^P$,

$$\tilde{\mathcal{V}}^P \equiv \left\{w\in H^1(\Omega):\,\mathcal{L}(w) := -\Delta w\in L^2(\Omega),\,\mathcal{B}(w) := \nabla w\cdot\mathbf{n}|_\Gamma\in L^2(\Gamma)\right\}$$

then the discrete solution u_h satisfies the following error estimate

$$|||u-u_h|||_P \leq \left(1+\frac{2}{1-\frac{1}{\gamma}}\right)\inf_{v_h\in\mathcal{V}_h^P}|||u-v_h|||_P,$$

where

$$|||v|||_P^2 \equiv \|\nabla w\|_{0,\Omega}^2 + \frac{h}{C_{\mathrm{tr}}}\|\nabla w\cdot\mathbf{n}\|_{0,\Gamma}^2 + 2\frac{C_{\mathrm{pen}}}{h}\|w\|_{0,\Gamma}^2.$$

Proof. Notice that Nitsche's method for the Poisson problem exactly fits into the abstract variational framework, as shown in Example 1.17. Hence the well-posedness and the error estimate follow directly from Theorem 1.16. □

Remark 1.12. It should be noted that to avoid an excessively large penalty parameter in the presence of small cut cells, Nitsche's method usually appears along with a stabilization technique, such as ghost penalty stabilization, which can refer to Burman et al. [2].

1.6 Numerical Experiments

We employ Nitsche's method to solve the Poisson equation on geometrically complex domains such as vortex shear of a circular disk, where traditional body-fitted mesh methods pose significant challenges. The initial domain configuration at $T = 0$ is circular, which subsequently deforms under the vortex velocity field:

$$\boldsymbol{v}(\boldsymbol{p}, t) = \cos\left(\frac{\pi t}{4}\right)\begin{bmatrix} \sin^2(\pi p_x)\sin(2\pi p_y) \\ -\sin^2(\pi p_y)\sin(2\pi p_x) \end{bmatrix},$$

until final time $T = 2$.

The spatial discretization employs fourth-order finite elements on a cut-cell grid. Figure 1.1 illustrates the mesh configuration at refinement level $h = 1/16$.

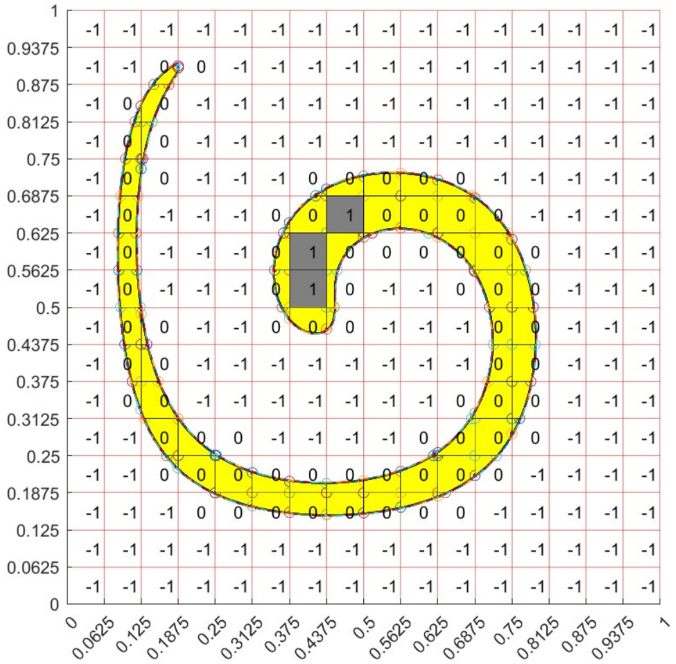

Figure 1.1: Cut-cell mesh topology with element classification at $h = 1/16$. Interface elements (Label 0) separate the active computational domain (Label 1) from exterior regions (Label -1).

Table 1.1 presents the numerical convergence results, demonstrating optimal convergence rates in both L^2 and L^∞ norms.

Table 1.1: Error norms and convergence rates for the cut-cell discretization

h	L^2 Error	Rate	L^∞ Error	Rate
1/16	2.2366×10^{-6}	—	6.1292×10^{-6}	—
1/32	1.3944×10^{-7}	4.00	4.7547×10^{-7}	3.69
1/64	6.9208×10^{-9}	4.33	2.6455×10^{-8}	4.17
1/128	3.4307×10^{-10}	4.33	1.8015×10^{-9}	3.88

[2] Burman, E., Claus, S., Hansbo, P., Larson, M. G., & Massing, A. (2014). CutFEM: Discretizing geometry and partial differential equations. International Journal for Numerical Methods in Engineering