

A REMARK ON THE "INF-SUP-CONDITION" IN THE FINITE ELEMENT METHOD

Joachim A. Nitsche

Dedicated to Professor Eugene Isaacson on his 70th Birthday

Abstract: The necessity of the "inf-sup-condition" used in Galerkin and hybrid/mixed methods is analyzed.

Key words: Galerkin, hybrid/mixed methods, finite elements

AMS(MOS) subject classifications: 65N30

1.

Let X, Y be reflexive Banach spaces and $a(\cdot, \cdot)$ be a bilinear form mapping $X \times Y$ onto \mathbb{R} , i.e. for any pair $(x, y) \in X \times Y$ a real $a(x, y) \in \mathbb{R}$ is associated and $a(\cdot, \cdot)$ is linear with respect to both arguments.

For fixed $x \in X$ resp. $y \in Y$ by $g(\cdot) = g_x(\cdot) = a(x, \cdot)$ resp. $f(\cdot) = f_y(\cdot) = a(\cdot, y)$ an element $g \in Y^*$ resp. $f \in X^*$ is given. In this way the bilinear form $a(\cdot, \cdot)$ defines a linear mapping $A : X \rightarrow Y^*$ resp. $B : Y \rightarrow X^*$. A and consequently B are bounded, i.e. $A \in \mathcal{B}(X, Y^*)$ and $B \in \mathcal{B}(Y, X^*)$, iff there exists a constant M with

$$(1.1) \quad a(x, y) \leq M \|x\| \|y\| \quad \text{for } x \in X, y \in Y.$$

The smallest constant M in (1.1) is the norm of A and B . Since X, Y are assumed to be reflexive A and B are their duals: $B = A^*, A = B^*$.

The operator A resp. B is injective, if the condition holds true

$$(1.2) \quad a(x, y) = 0 \quad \text{for } y \in Y \text{ implies } x = 0$$

resp.

$$(1.3) \quad a(x, y) = 0 \quad \text{for } x \in X \text{ implies } y = 0.$$

On the contrary (1.2) resp. (1.3) implies that the range of B resp. of A is dense in X^* resp. in Y^* .

Let (1.2), (1.3) be valid. Necessary and sufficient that A^{-1} is a bounded mapping of Y^* onto X is the condition: There exists a positive constant m such that

$$(1.4) \quad m \|x\| = \sup \{ a(x, y) \mid y \in Y \wedge \|y\| = 1 \}$$

holds true.

Remark 1.1:

Since the best constant m in (1.4) may be characterized by

$$(1.5) \quad m = \inf_{x \in X} \sup_{y \in Y} \{ a(x, y) \mid \|x\| = \|y\| = 1 \}$$

the condition (1.4) is often called "inf-sup-condition".

The best constant m in (1.4) is related to the norm of A^{-1} by $\|A^{-1}\| = m^{-1}$. Since B is the operator dual to A and $B^{-1} = (A^{-1})^*$ a consequence of (1.4) is

$$(1.6) \quad m \|y\| \leq \sup \{ a(x, y) \mid x \in X \wedge \|x\| = 1 \}$$

valid for all $y \in Y$.

The above formulation is due to Nečas, J. (1962) and Nirenberg, L. (1955).

2.

Let $a(\cdot, \cdot)$ be a bilinear form as discussed in section 1. The Galerkin procedure provides a method to get approximations on the solution $x \in X$ of

$$(2.1) \quad a(x, y) = g(y) \quad \text{for } y \in Y$$

Let $\{S_n, T_n\}$ be a sequence of finite dimensional approximation and trial spaces, i.e. $S_n \subset X, T_n \subset Y$. The Galerkin approximations $x_n = G_n x \in S_n$ on the solution x of (2.1) are defined by

$$(2.2) \quad a(x_n, \psi) = g(\psi) \quad \text{for } \psi \in T_n.$$

Equivalently we may rewrite (2.2) in the form

$$(2.2') \quad a(x_n, \psi) = a(x, \psi) \quad \text{for } \psi \in T_n$$

We cite the essential convergence result due to Babuška, I. (1971): The Galerkin procedure is almost best approximating in the sense

$$(2.3) \quad \|x - x_n\| \leq C \inf \{ \|x - \varphi\| \mid \varphi \in S_n \}$$

with C being independent of n if (1.3), (1.4) hold true with X, Y replaced by S_n, T_n and a positive constant \underline{m} in (1.4) independent of n.

3.

In the applications quite often the spaces X, Y coincide and are the cartesian product of two Hilbert spaces: $X = Y = \tilde{H} \times \tilde{H}$. In addition the bilinear form $a(\cdot, \cdot)$ has the structure

$$(3.1) \quad a(x, y) = \alpha(\tilde{x}, \tilde{y}) + \beta(\tilde{x}, \tilde{y}) + \beta(\tilde{y}, \tilde{x})$$

with $x = (\tilde{x}, \tilde{x})$ and $y = (\tilde{y}, \tilde{y})$. Typically $\alpha(\cdot, \cdot)$ is a bounded, symmetric, and positive definite bilinear form on $\tilde{H} \times \tilde{H}$. In this situation the choice $S_n = T_n$ is usual with $S_n = \tilde{S}_n \times \tilde{S}_n \subset \tilde{H} \times \tilde{H}$. The result of Brezzi, F. (1975) is: Let $x_n = (\tilde{x}_n, \tilde{x}_n)$ be the Galerkin approximation on $x = (\tilde{x}, \tilde{x})$ according to the variational equation (2.2). If the bilinear form $\beta(\cdot, \cdot)$ is bounded and admits the condition

$$(3.2) \quad \underline{m} \|\tilde{f}\| \leq \sup \{ \beta(\tilde{u}, \tilde{f}) \mid \tilde{u} \in \tilde{S}_n \wedge \|\tilde{u}\| = 1 \}$$

for all $\tilde{f} \in \tilde{S}_n$ with $\underline{m} > 0$ independent of n then x_n is almost best approximating in the sense (with C independent of n):

$$(3.3) \quad \begin{aligned} & \| \tilde{x} - \tilde{x}_n \| + \| \tilde{x} - \tilde{x}_n \| \\ & \leq C \inf \{ \| \tilde{x} - \tilde{f} \| + \| \tilde{x} - \tilde{f} \| \mid \tilde{f} \in \tilde{S}_n \wedge \|\tilde{f}\| = 1 \} \end{aligned}$$

4.

Let H be a separable Hilbert space and A a positive definite, self-adjoint operator in H. The bilinear form

$$(4.1) \quad a(x, y) = (Ax, y)$$

defines a scalar product $(\cdot, \cdot)_A$ in $D(A)$, the Hilbert space H_A is defined by completion of $D(A)$ with respect to the corresponding 1-norm. Correspondingly the Hilbert space H_2 is $D(A)$ equipped with the graph norm or equivalently with $\|\cdot\|_2 = \|A \cdot\|$.

In case of $T_n = S_n$ and $S_n \subset H_1$ by (2.2) the Ritz approximation $x_n = P_n x$ on x is defined, it is the orthogonal projection of x onto S_n in the metric of H_1 .

The inequalities stated in section 1 are valid for the bilinear form (4.1) with $X = H_2, Y = H$. Thus the Ritz procedure may be viewed as Galerkin procedure for this choice of spaces, of course $S_n \subset H_2$ is necessary. The convergence $x_n = P_n x \rightarrow x$ in $X = H_2$ is equivalent to the convergence of the residuum $AR_n x - Ax = AP_n x - g \rightarrow 0$ in the norm of the underlying Hilbert space H. It was shown by Polskii, I. (1962): Let the sequence $\{S_n\}$ be dense in H_2 . If the condition

$$(4.2) \quad \inf \{ \|P_n A \tilde{f}\| \mid \tilde{f} \in S_n \wedge \|A \tilde{f}\| = 1 \} \geq \underline{g} > 0$$

holds true with P_n denoting the orthogonal projection of H onto S_n then the residuum converges to zero: $AP_n x - g \rightarrow 0$ in H.

The condition (4.2) may be reformulated as an inf-sup-condition. Because of $P_n A \tilde{f} \in S_n$ it is

$$(4.3) \quad \|P_n A \tilde{f}\| = \sup \{ (P_n A \tilde{f}, \eta) \mid \eta \in S_n \wedge \|\eta\| = 1 \}$$

and further because of

$$(4.4) \quad \begin{aligned} (P_n A \tilde{f}, \eta) &= (A \tilde{f}, \eta) \\ &= a(\tilde{f}, \eta) \end{aligned}$$

condition (4.2) is equivalent to

$$(4.5) \quad \underline{g} \|\tilde{f}\|_{H_2} \leq \sup \{ a(\tilde{f}, \eta) \mid \eta \in S_n \wedge \|\eta\| = 1 \}.$$

For the sake of completeness we mention the two earlier papers: Polskii (1952) and (1956) in which sufficient conditions for the convergence of Galerkin procedures are given.

5.

The conditions stated in sections 2. - 4. are primarily sufficient conditions. Polskii, N.I. (1962) mentions that the condition (4.2) resp. a corresponding condition is necessary. With respect to this Polskii refers to Krasnoseiskii (1950). But in this paper only the hint is given: "It can be shown ..." (see lines 19 to 21 on page 1123 of the paper cited). The aim of this paper is to show the necessity of (1.4) in order to have an error estimate of the type (2.3) for the Galerkin method (2.2).

If $N_h(S)$, $N_h(T)$ denote the dimensions of S_h , T_h then the variational equation (2.2) amounts to a system of $N_h(T)$ linear equations for $N_h(S)$ unknown. The Galerkin approximations are well defined if $N_h(T) = N_h(S)$ (for simplicity let n indicate the dimension) and furthermore, if the matrix of the corresponding linear system has maximal rank. This condition may be formulated in the way:

Let $\xi \in S_h$ resp. $\eta \in T_h$ be fixed. Then

$$(5.1) \quad \begin{aligned} a(\xi, \psi) &= 0 && \text{for } \psi \in T_h \text{ imply } \xi = 0, \\ a(\varphi, \eta) &= 0 && \text{for } \varphi \in S_h \text{ imply } \eta = 0. \end{aligned}$$

Different from the situation discussed in section 1 the above two conditions are equivalent.

To any $x \in X$ the corresponding solution of (2.2) is denoted by $x_h = G_h x$. The map - ping $G_h : X \rightarrow S_h$ is a linear projection: If x happens to be an element of S_h then obviously $G_h x$ is identical with x . Equivalent to the error estimate (2.3) is the uniform boundedness of the operators G_h

$$(5.2) \quad \|G_h\| \leq c$$

The necessary condition such that (5.2) holds true is

Proposition 5.1:

The pairs (S_h, T_h) share a common inf-sup-condition in the sense that there exists a bound $\alpha > 0$ with the property that

$$(5.3) \quad \alpha \|\varphi\| \leq \sup \{ a(\varphi, \psi) \mid \psi \in T_h \wedge \|\psi\| = 1 \}$$

holds true for $\varphi \in S_h$ and $n \in \mathbb{N}$.

Proof:

We will work with special bases in S_h and T_h . Of course the projection operators G_h are independent of these special choices. Because of (5.1) there exist bases $\{\varphi_i\}$, $\{\psi_k\}$ of $S = S_h$, $T = T_h$ according to

$$(5.4) \quad a(\varphi_i, \psi_k) = \delta_{ik} \quad \text{for } i, k = 1, \dots, n$$

We introduce $\Gamma = A(S) \subset Y^*$ spanned by

$$(5.5) \quad \chi_i = A\varphi_i \quad \text{for } i = 1, \dots, n$$

Then $\{\chi_i\}$, $\{\psi_k\}$ form a dual pair of bases in Γ , T according to

$$(5.6) \quad \chi_i(\psi_k) = \delta_{ik} \quad \text{for } i, k = 1, \dots, n$$

Let the spaces $N \subset Y$, $M \subset Y^*$ be defined by

$$(5.7) \quad \begin{aligned} N &= \{ y \mid y \in Y \wedge \chi(y) = 0 \text{ for } \chi \in \Gamma \}, \\ M &= \{ g \mid g \in Y^* \wedge g(\eta) = 0 \text{ for } \eta \in T \}. \end{aligned}$$

Then Y , Y^* are the direct sums

$$(5.8) \quad \begin{aligned} Y &= T \oplus N \\ Y^* &= \Gamma \oplus M \end{aligned}$$

After these preparations we start analyzing condition (5.2). We remark: for $x \in X$ given the element $g = Ax \in Y^*$ admits the decomposition

$$(5.9) \quad g = Ax = \chi - \mu$$

with $\chi \in \Gamma$, $\mu \in M$. Because of the choice of the bases in S , T we have

$$(5.10) \quad AGx = \chi$$

From the following sequence of estimates

$$\begin{aligned}
 \|\chi\| &\leq \|A\|Gx\| \\
 (5.11) \quad &\leq c\|A\|\|x\| \\
 &\leq c\|A^{-1}\|\|A\|\|g\|
 \end{aligned}$$

we derive

$$(5.12) \quad \|\chi\| \leq c\|\chi-\mu\|$$

resp.

$$(5.13) \quad \|\chi\| \leq c \inf\{\|\chi-\mu\| \mid \mu \in M\}$$

Now let $\eta \in T$ be fixed. There exists a $g \in Y^*$ with

$$\begin{aligned}
 (5.14) \quad \|g\| &= 1 \\
 g(\eta) &= \|\eta\|
 \end{aligned}$$

The functional g admits the decomposition

$$(5.15) \quad g = \chi - \mu$$

with $\chi \in \Gamma, \mu \in M$. The consequence is

$$(5.16) \quad g(\eta) = \chi(\eta)$$

Further we have because of (5.13) and (5.15)

$$(5.17) \quad \|g\| \geq c^{-1}\|\chi\|$$

and therefore because of (5.14) and (5.16)

$$\begin{aligned}
 (5.18) \quad \|\eta\| &= \chi(\eta) \\
 &\leq c\|\chi(\eta)\| / \|\chi\|
 \end{aligned}$$

Let $\xi \in S$ be defined by $\xi = A^{-1}\chi$. Then we get furthermore

$$(5.19) \quad \|\eta\| \leq c\alpha(\xi, \eta)\|A^{-1}\| / \|\xi\|$$

and finally

$$(5.20) \quad \|\eta\| \leq c \sup_{\alpha(\xi, \eta) / \|\xi\|}$$

This shows that necessarily an inf-sup-condition of the above type has to hold true. This result may be summarized by

Lemma 5.2:

Within the setting discussed in this section the inf-sup-condition

$$(5.21) \quad c\|\eta\| \leq \sup\{\alpha(\xi, \eta) \mid \xi \in S_n \wedge \|\xi\| = 1\}$$

valid for $\eta \in \Gamma_n$ is a consequence of (5.2).

The inequality (5.13) was the key to derive (5.20) resp. (5.21). A consequence of (5.13) is a corresponding estimate for the elements $\eta \in T = T_n$. In order to show this let $\eta \in T$ be fixed with $\|\eta\| = 1$. Further let g, χ, μ be defined according to (5.14, 15). Because of (5.17) we conclude

$$(5.22) \quad \|\chi\| \leq c$$

Now let $v \in N$ be arbitrary. From (5.14) and the definition of the subspace N (5.7) we derive

$$\begin{aligned}
 (5.23) \quad 1 &= g(\eta) \\
 &= \chi(\eta-v) \\
 &\leq \|\chi\|\|\eta-v\|
 \end{aligned}$$

In view of (5.13, 24) we get for $\eta \in T$ with $\|\eta\| = 1$

$$(5.24) \quad \inf\{\|\eta-v\| \mid v \in N\} \geq c$$

(5.24) is the counterpart of (5.13) for the elements η of T . Following the arguments from (5.13) to (5.20) we conclude that a corresponding inf-sup-condition holds true with S, T replaced by T, S . This result may be interpreted in the following way: Let us consider the problem (dual to (2.1))

$$(5.25) \quad a(x, y) = f(x) \quad \text{for } x \in X$$

for a given $f \in X^*$ and the corresponding Galerkin approximation $y_h \in T_n$ on $y = B^{-1}f$ defined by

$$(5.26) \quad a(\varphi, y_h) = a(\varphi, y) \quad \text{for } \varphi \in S_h$$

Then y_h is almost best approximating in the sense

$$(5.27) \quad \|y - y_h\| \leq C \inf \{ \|y - \psi\| \mid \psi \in T_n \}$$

6.

In the applications mostly the spaces $\{S_n\}, \{T_n\}$ are nested, i.e. the inclusions hold true

$$(6.1) \quad S_n \subset S_{n+1}, \quad T_n \subset T_{n+1}$$

Then for $x \in X$ respective $y \in Y$ the quantities

$$(6.2) \quad \begin{aligned} d_n(x) &= \inf \{ \|x - \varphi\| \mid \varphi \in S_n \} \\ d_n(y) &= \inf \{ \|y - \psi\| \mid \psi \in T_n \} \end{aligned}$$

are non-increasing. The sequence $\{S_n\}$ respective $\{T_n\}$ is called dense in X respective Y if $d_n(x)$ respective $d_n(y)$ tends to zero for $n \rightarrow \infty$. A consequence of the derivations of section 5 is

Theorem 6.1:

Let $\{S_n, T_n\}$ fulfill Proposition 5.1. If in addition $\{S_n\}$ is dense in X then also $\{T_n\}$ is dense in Y .

Remark 6.2:

Obviously the role of X, Y respective S_n, T_n may be interchanged.

Proof:

If $\{T_n\}$ is not dense in Y , i. e. if

$$(6.3) \quad Y_0 = \bigcup_{n=1,2,\dots} T_n$$

is different from Y , there would exist an element $\tilde{y} \in Y$ with $\tilde{y} \notin Y_0$. By the Hahn Banach Theorem there would exist an element $g \in Y^*$ with

$$(6.4) \quad \begin{aligned} g(y) &= 0 & \text{for } y \in Y_0 \\ g(\tilde{y}) &= 1 \end{aligned}$$

The condition that $\{S_n\}$ is dense in X is equivalent that $\{\Gamma_n\}$ - see (5.5) - is dense in Y^* . Therefore to a given $\epsilon > 0$ there exists an $n = n_\epsilon$ and an element $\chi_n \in \Gamma_n$ according to

$$(6.5) \quad \|g - \chi_n\| \leq \epsilon$$

The inf-sup-condition (5.3) implies

$$(6.6) \quad \|\chi_n\| \leq c \sup \{ \chi_n(\psi) \mid \psi \in T_n \wedge \|\psi\| = 1 \}$$

with some constant c independent of n . Because of

$$(6.7) \quad g(\psi) = 0$$

we get

$$(6.8) \quad \chi_n(\psi) \leq \epsilon \quad \text{for } \psi \in T_n \wedge \|\psi\| = 1$$

and thus because of (6.6)

$$(6.9) \quad \|\chi_n\| \leq c \epsilon$$

This contradicts (6.5) and $g \neq 0$ if ϵ is chosen sufficiently small.

Bibliography:

Babuska, I. (1971)
 Error bounds for finite element method
 Numer. Math. 16, 322-339 (1971)

Brezzi, F. (1975)
 Sur la methode des elements finis hybrides pour le probleme biharmonique
 Numer. Math. 24, 103-131 (1975)

Krasnoselskii, M.A. (1950)
 Convergence of Galerkin's method for nonlinear equations
 Dokl. Akad. Nauk SSSR 73, 1121-1124 (1950)

Necas, J. (1962)

Sur une méthode pour résoudre les équations dérivées partielles du type elliptique voisine de la variationnelle
Ann. Sc. Norm. Sup., Pisa, Ser III, 15, 305-326 (1962)

Nirenberg, L. (1955)

Remarks on strongly elliptic partial differential equations
Comm. Pure App. Math. 8, 649-675 (1955)

Polskii, N.I. (1952)

Some generalizations of B.G. Galerkin's method
Dokl. Akad. Nauk SSSR, 85, 469-472 (1952)

----- (1956)

On a general scheme of application of approximation methods
Dokl. Akad. Nauk SSSR, 111, 1181-1184 (1956)

----- (1962)

Projective methods in applied mathematics
Soviet Math. Dokl. 3, 488-492 (1962).