

Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes

Erik Burman ^a, Alexandre Ern ^b

^a DMA, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
E-mail: Erik.Burman@epfl.ch

^b CERMICS, Ecole nationale des ponts et chaussées, 77455 Marne-la-Vallée cedex 2, France
E-mail: ern@cermics.enpc.fr

(Reçu le jour mois année, accepté après révision le jour mois année)

Abstract. We derive a nonlinear stabilized Galerkin approximation of the Laplace operator for which we prove a discrete maximum principle on arbitrary meshes and for arbitrary space dimension without resorting to the well-known acute condition or generalizations thereof. We also prove the existence of a discrete solution and discuss the extension of the scheme to convection-diffusion-reaction equations. Finally, we present examples showing that the new scheme cures local minima produced by the standard Galerkin approach while maintaining first-order accuracy in the H^1 -norm. © 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Principe du maximum discret pour des approximations de Galerkin du Laplacien sur des maillages quelconques

Résumé. Nous introduisons un terme de stabilisation non-linéaire pour lequel nous prouvons un principe du maximum discret pour des approximations de type Galerkin du Laplacien. Le principe du maximum discret est satisfait en dimension quelconque et sans hypothèse particulière sur le maillage. On s'affranchit notamment de la condition bien connue d'acuité ou des généralisations de celle-ci. Nous prouvons également l'existence d'une solution discrète et proposons une extension du schéma aux équations de convection-diffusion-réaction. Enfin, nous présentons des résultats numériques montrant que le schéma élimine bien les minima locaux produits par la méthode de Galerkin standard tout en maintenant une convergence à l'ordre un en norme H^1 . © 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Version française abrégée

La conception de méthodes d'approximation robustes pour les équations aux dérivées partielles (EDP) conjugue plusieurs critères. Dans de nombreuses applications, l'un de ceux-ci est de garantir que la solution

Note présentée par Philippe G. CIARLET

S0764-4442(00)0????-?/FLA

© 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

approchée satisfasse un principe du maximum discret (PMD). Par exemple, dans la simulation du transport d'espèces chimiques par des écoulements inertes ou réactifs, on souhaite que les concentrations restent positives. Des schémas garantissant un PMD pour des équations de convection-diffusion-réaction ont été étudiés récemment dans [1, 2]. Ces schémas font intervenir une approximation de type Galerkin standard par éléments finis H^1 -conformes et un terme de stabilisation non-linéaire (parfois appelé terme de capture de choc). Dans [1], la stabilisation est assurée par des termes de diffusion à la fois le long des lignes de courant et orthogonalement à celles-ci. Dans [2], le schéma combine des dérivées tangentielles le long des arêtes du maillage et le saut des dérivées normales aux interfaces.

Les schémas obtenus dans [1, 2] garantissent rigoureusement un PMD sous une hypothèse géométrique sur le maillage, typiquement une condition de faible acuité. Cette hypothèse est nécessaire afin de contrôler les termes d'ordre supérieur dans l'EDP, i.e., ceux dus à la diffusion. En effet, il est bien connu [3] que l'opérateur de Laplace discrétisé avec des éléments finis continus et linéaires par morceaux sur des maillages faiblement aigus conduit à une M-matrice et donc à un PMD. Cette condition géométrique peut être relaxée en considérant par exemple des triangulations de type Delaunay en deux dimensions ; des extensions en trois dimensions ont été étudiées dans [4, 5]. Toutefois, à ce jour, toutes les discrétisations du Laplacien qui garantissent un PMD mettent en jeu des contraintes géométriques sur le maillage. L'objectif de cette Note est l'analyse (en dimension quelconque) d'un nouveau schéma numérique garantissant un PMD pour le Laplacien sur des maillages quelconques. Nous considérons des approximations par éléments finis continus et linéaires par morceaux car ceux-ci interviennent très souvent dans les codes utilisés par les ingénieurs.

On considère le problème (1) avec des données $f \in L^2(\Omega)$ et $g \in H^{1/2}(\partial\Omega)$, Ω étant un ouvert borné et connexe de \mathbb{R}^d . Soit \mathcal{T} un maillage de Ω par des simplexes. Soit \mathcal{F} l'ensemble des faces intérieures du maillage. Pour $F \in \mathcal{F}$, on note $\mathcal{E}(F)$ l'ensemble des arêtes de F . Pour une arête e , on note h_e sa longueur et t_e un vecteur unitaire porté par celle-ci. On considère le problème discret (4), l'espace V_h^g étant défini par (2). La forme semi-linéaire $a(U; v)$ comprend le terme issu de la méthode de Galerkin standard et un terme de stabilisation non-linéaire $j(U; v)$ donné par (5). Celui-ci fait intervenir, via $\psi_F(U; v)$, les dérivées tangentielles de la solution discrète et de la fonction test le long des arêtes du maillage. Le théorème 2.1 assure l'existence d'une solution au problème non-linéaire (4). La preuve consiste à introduire l'opérateur T_ϵ qui à $\tilde{U} \in V_h^0$ associe la solution unique, U , de (6). On utilise le théorème de Brouwer pour montrer l'existence d'un point fixe, U_ϵ , de T_ϵ , puis on extrait une sous-suite convergeant vers une solution de (4).

Le résultat principal de cette Note est donné par le théorème 2.2 : si le coefficient $\delta(U)$ est évalué sur chaque face F selon (7) où $[\nabla U]_F$ est le saut de ∇U à-travers la face et si la constante c dans (5) est telle que $c > \frac{1}{d(d-1)}$, alors le schéma (4) jouit du PMD suivant : si $f \geq 0$, toute solution de (4) atteint son minimum au bord de Ω . La preuve consiste à raisonner par l'absurde. On suppose que U atteint son minimum en un sommet S_i situé à l'intérieur du maillage. Soit Ω_i le macro-élément formé des simplexes contenant S_i . En testant avec la fonction de base nodale associée à S_i , on montre que $(f, w_i)_\Omega = (\nabla U, \nabla w_i)_\Omega + j(U; w_i)$ est majoré par $\sum_{F \in \mathcal{F}(S_i)} \alpha_F |\nabla U|_F$ où $\mathcal{F}(S_i)$ est l'ensemble des faces contenant S_i et où tous les coefficients α_F sont négatifs. Comme $f \geq 0$, on en déduit que ∇U est constant sur Ω_i . Enfin, dans la Section 4, nous présentons quelques résultats numériques. Nous considérons des maillages sur lesquels l'approximation de Galerkin standard du Laplacien conduit à des violations du principe du maximum. La stabilisation numérique proposée permet d'éliminer ces violations tout en conservant un degré de précision équivalent à l'approximation standard (erreur d'ordre un en norme H^1).

En conclusion, on notera que même si le schéma proposé présente le défaut de transformer un problème linéaire en un problème non-linéaire, il a vocation à être utilisé pour des problèmes plus complexes, mettant déjà en jeu des systèmes d'EDP couplées non-linéaires. En particulier, l'opérateur (5) peut être employé conjointement à celui introduit dans [2] afin d'obtenir des approximations de l'équation de convection-diffusion-réaction (8) satisfaisant un PMD. Le schéma numérique est donné par (9) et le terme de stabilisation non-linéaire par (10).

1. Introduction

In several applications, one important criterion to design robust approximation methods to partial differential equations (PDEs) is to guarantee that the discrete solution satisfies some type of maximum principle. For instance, in problems where chemical species are transported by inert or reactive flows, species concentrations should remain non-negative. Numerical schemes ensuring a discrete maximum principle (DMP) for convection-diffusion-reaction equations have been investigated recently in [1, 2]. These schemes are based on a standard Galerkin formulation with H^1 -conforming finite elements supplemented by a nonlinear stabilization term (sometimes termed a shock capturing term). In [1], the stabilization relies on solution-dependent weightings of streamline and crosswind diffusion. In [2], the scheme uses tangential derivatives along element edges and the jumps of normal derivatives across interfaces.

A DMP can be established rigorously for the schemes in [1, 2] provided the mesh satisfies some type of geometric condition, typically a weakly acute condition. This assumption is needed to control the higher-order terms in the PDE, i.e., those associated with diffusion. In particular, one uses the well-known fact that the Laplace operator discretized using continuous, piecewise linear finite elements satisfies a DMP on weakly acute meshes [3]. In two dimensions, this geometrical condition can be extended to Delaunay meshes; extended conditions can also be designed in three dimensions [4, 5]. The current situation is that all the discretization schemes for the Laplace operator that are endowed with a maximum principle must comply with some type of geometrical constraint on the mesh. These conditions are sometimes difficult to satisfy in practice, especially in three dimensions.

The purpose of this work is to analyze a stabilized Galerkin approximation of the Laplacian guaranteeing a DMP on arbitrary meshes. The main motivation for introducing this scheme is not to solve approximately Laplace-type equations, but to use it in the context of nonlinear PDEs coupling reactive, convective, and diffusive effects. In this case, the additional complexity brought in by the nonlinear stabilization is reasonable. We focus on continuous, piecewise linear finite elements since they are often used in engineering codes.

2. Main results

Consider the problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases} \quad (1)$$

where Ω is an open, bounded, connected subset of \mathbb{R}^d with a Lipschitz boundary $\partial\Omega$, d is the (arbitrary) space dimension, $f \in L^2(\Omega)$, and $g \in H^{1/2}(\partial\Omega)$. It is well-known that Problem (1) satisfies the following maximum principle: if $f \geq 0$, then $\inf_{\Omega} u \geq \inf_{\partial\Omega} g$; see, e.g., [6, p.179].

2.1. The discrete setting

Let \mathcal{T} be a simplicial mesh of Ω . For the sake of simplicity, we assume that \mathcal{T} covers $\bar{\Omega}$ exactly. Let

$$V_h^g = \{ v \in C^0(\bar{\Omega}); \forall T \in \mathcal{T}, v|_T \in P_1(T); v = \mathcal{P}g \text{ on } \partial\Omega \}, \quad (2)$$

where \mathcal{P} denotes the L^2 -projection onto the space of continuous, piecewise affine functions on the boundary (the Lagrange interpolate could have been considered as well). Let $(u, v)_{\Omega}$ denote the $L^2(\Omega)$ -scalar product and $\|u\|_{\Omega} = (u, u)_{\Omega}^{1/2}$ the associated norm. The standard Galerkin approximation to (1) is as follows: Find $\tilde{U} \in V_h^g$ such that

$$(\nabla \tilde{U}, \nabla v)_{\Omega} = (f, v)_{\Omega}, \quad \forall v \in V_h^0. \quad (3)$$

In one dimension, the associated stiffness matrix is an M-matrix regardless of the mesh, so that a DMP always holds. Therefore, we shall henceforth treat the case $d \geq 2$. In this case, the stiffness matrix is an

M-matrix only if the mesh complies with some geometrical constraints, such as those derived in [3, 4, 5]. Furthermore, DMP violations can be observed if these conditions are not satisfied (see Section 4 for a counter-example).

We adopt the terminology of the three-dimensional case, being understood that in two dimensions, a face is also an edge. Let \mathcal{F} denote the set of interior faces of the mesh and \mathcal{E} the set of interior edges. For $F \in \mathcal{F}$ and $e \in \mathcal{E}$, we denote by h_F and h_e their respective diameter and by $|F|$ the $(d-1)$ -measure of F . Let t_e be a unit vector along e (its orientation is irrelevant). For a face $F \in \mathcal{F}$, $\mathcal{E}(F)$ denotes the set of edges it contains. Let S_i be an interior vertex in the mesh, let $w_i \in V_h^0$ be the associated nodal basis function, and denote by Ω_i the macro-element formed by the elements in \mathcal{T} sharing the vertex S_i . Let $\mathcal{F}(S_i)$ (resp., $\mathcal{E}(S_i)$) denote the set of faces (resp., edges) containing S_i . For $U \in V_h^g$ and an interior face $F = T_1 \cap T_2$, where T_1 and T_2 are two distinct elements of \mathcal{T} with respective outer normals n_1 and n_2 , introduce the jump $[\nabla U]_F = \nabla(U|_{T_1}) \cdot n_1 + \nabla(U|_{T_2}) \cdot n_2$.

To circumvent any geometrical condition on the mesh, we introduce a nonlinear stabilization term involving tangential derivatives along element edges and the jumps of the gradient across interfaces. The stabilized Galerkin approximation to (1) is: Find $U \in V_h^g$ such that

$$a(U; v) := (\nabla U, \nabla v)_\Omega + j(U; v) = (f, v)_\Omega, \quad \forall v \in V_h^0, \quad (4)$$

with

$$j(U; v) = c \sum_{F \in \mathcal{F}} \delta(U)|_F \psi_F(U; v), \quad (5)$$

where c is a constant only depending on the space dimension, δ is a function of U defined on each face, and $\psi_F(U; v) = \sum_{e \in \mathcal{E}(F)} h_e \operatorname{sign}(\nabla U \cdot t_e) \nabla v \cdot t_e$.

2.2. Existence of a discrete solution

THEOREM 2.1. – *The nonlinear problem (4) admits at least one solution.*

Proof. – The proof is similar to the one presented in [2] for convection-diffusion-reaction equations and is, therefore, only sketched. Consider for simplicity $g = 0$. Let $\epsilon > 0$ and consider the mapping $T_\epsilon : V_h^0 \ni \tilde{U} \rightarrow U \in V_h^0$ where U solves

$$(\nabla U, \nabla v)_\Omega + c \sum_{F \in \mathcal{F}} \delta(\tilde{U})|_F \left(\sum_{e \in \mathcal{E}(F)} h_e \frac{(\nabla U \cdot t_e)(\nabla v \cdot t_e)}{|\nabla \tilde{U} \cdot t_e| + \epsilon} \right) = (f, v)_\Omega, \quad \forall v \in V_h^0. \quad (6)$$

The application T_ϵ is well-defined owing to the Lax-Milgram lemma. To prove that T_ϵ has at least one fixed point, say U_ϵ , one first deduces from (6) an a priori estimate on $\|\nabla U\|_\Omega$. The continuity of T_ϵ in V_h^0 equipped with the H^1 -norm is then easily established. Existence of a fixed point U_ϵ subsequently results from Brouwer's Theorem. Since the sequence $\{U_\epsilon\}_\epsilon$ is in a finite-dimensional ball, we may extract a subsequence, still denoted by $\{U_\epsilon\}_\epsilon$, such that $U_\epsilon \rightarrow U \in V_h^0$ in $H_0^1(\Omega)$ as $\epsilon \rightarrow 0$. By passing to the limit in (6), we conclude that U satisfies (4). \square

2.3. Discrete maximum principle

THEOREM 2.2. – *Assume that $\delta(U)$ in (5) is given by*

$$\forall F \in \mathcal{F}, \quad \delta(U)|_F = |F| |\nabla U|_F, \quad (7)$$

and that the constant c is such that $c > \frac{1}{d(d-1)}$. Let U be a solution to (4). Then the following DMP holds for (4): if $f \geq 0$, U reaches its minimum on the boundary.

Discrete maximum principle for the Laplacian on arbitrary meshes

Proof. – Let S_i be an interior vertex and assume that U reaches a local minimum over the macro-element Ω_i at S_i . Taking $v = w_i$ in (4) and integrating by parts the standard Galerkin term yields

$$\sum_{F \in \mathcal{F}(S_i)} ([\nabla U]_F, w_i)_F + c \sum_{F \in \mathcal{F}(S_i)} \delta(U)|_F \psi_F(U; v) = (f, w_i)_\Omega.$$

Because continuous, piecewise linear finite elements are used, $[\nabla U]_F$ is constant on F . Furthermore, for $F \in \mathcal{F}(S_i)$, $\int_F w_i = \frac{1}{d}|F|$. Moreover, for $e \in \mathcal{E}(F)$, $\nabla w_i \cdot t_e = 0$ if $e \notin \mathcal{E}(S_i)$ while $h_e \text{sign}(\nabla U \cdot t_e) \nabla w_i \cdot t_e \leq -1$ otherwise. As a result, we infer

$$(f, w_i)_\Omega \leq \sum_{F \in \mathcal{F}(S_i)} \delta(U)|_F \left(\frac{1}{d} - c(d-1) \right),$$

since there are $(d-1)$ edges in $\mathcal{E}(F) \cap \mathcal{E}(S_i)$. Assume $f \geq 0$. Then, the left-hand side is non-negative by assumption while the right-hand side is non-positive owing to the assumption on c . Therefore, $\delta(U)|_F = 0$ for all $F \in \mathcal{F}(S_i)$. Hence, ∇U is constant over the macro-element Ω_i , and we conclude easily. \square

Remark 1. – The fact that the constant c scales as $1/d^2$ indicates that the higher the space dimension, the harder it is for the standard Galerkin approximation to violate the DMP.

Remark 2. – The above proof shows that all the faces must not be included in the stabilization operator to guarantee a DMP. First, notice that the positive part of $[\nabla U]_F$ instead of its absolute value can be used in (7) since only the faces for which $[\nabla U]_F > 0$ need stabilization. Furthermore, advantage can also be drawn from the fact that the stabilization operator acts locally. For instance, in two dimensions, define η_F to be zero if the mesh is locally Delaunay in all the macro-elements containing F and to be one otherwise. Then, a DMP holds if $\delta(U)|_F = \eta_F |F| |\nabla U|_F$.

2.4. Extension to convection-diffusion-reaction problems

The discrete Laplace operator in (4) can be used in conjunction with the stabilized schemes derived in [2] for convection-diffusion-reaction equations. Consider the problem

$$\begin{cases} \beta \cdot \nabla u + \sigma u - \varepsilon \Delta u = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases} \quad (8)$$

with $\beta \in \mathbb{R}^d$, $\sigma \geq 0$, and $\varepsilon > 0$. One reasonable assumption on the mesh is necessary to establish a DMP, namely that there exists a constant ρ , independent of \mathcal{T} , such that for all $S_i \in \mathcal{T}$, $\max_{e \subset \Omega_i} h_e \leq \rho \min_{e \subset \Omega_i} h_e$. The stabilized Galerkin approximation to (8) is: Find $U \in V_h^g$ such that

$$(\varepsilon \nabla U, \nabla v)_\Omega + (\beta \cdot \nabla U, v)_\Omega + (\sigma U, v)_\Omega + j_{\varepsilon\beta\sigma}(U; v) = (f, v)_\Omega, \quad \forall v \in V_h^0, \quad (9)$$

with

$$\begin{cases} j_{\varepsilon\beta\sigma}(U; v) = c_{\rho\beta\sigma} \sum_{F \in \mathcal{F}} \delta_{\varepsilon\beta\sigma}(U)|_F \psi_F(U; v), \\ \delta_{\varepsilon\beta\sigma}(U) = (\varepsilon + |\beta| h_F + \sigma h_F^2) |[\nabla U]_F|. \end{cases} \quad (10)$$

Then, it can be shown using the results of [2] that if $f \geq 0$, $g \geq 0$, and the constant $c_{\rho\beta\sigma}$ is large enough, a discrete solution U to (9) is such that $U \geq 0$.

3. Numerical results

The purpose of this section is twofold: First, to illustrate the fact that the standard Galerkin method can fail to preserve the DMP and that this difficulty can be cured with the present scheme. Second, to assess the convergence order of the stabilized solution to (4) in the H^1 -norm.

Consider Equation (1) in the domain $\Omega = [0, 1] \times [0, 0.3]$ and set $f(x, y) = 1$ if $(x, y) \in [0, 0.5] \times [0, 0.075]$ and 0 otherwise. Take $g = 0$. Mesh Ω by splitting its edges into four cells and dividing the resulting sixteen quadrangular cells into four triangles along the two diagonals. The standard Galerkin solution yields an undershoot of 2% of the max-norm on this coarse mesh. Use now the stabilized Galerkin approximation (4). The nonlinear system of discrete equations is solved approximately using a damped Newton's method combined with a preconditioned Krylov linear solver (typically BiCGStab and ILU). The convergence behavior of Newton's method is significantly improved if one regularizes the sign operator in the stabilization operator, replacing it by $\text{sign}_\eta(x) = \tanh(x/\eta)$. We take $\eta = 10$, a choice for which Newton's method remains well-behaved. The stability properties of the scheme are not affected by this regularization since we obtain $U \geq 0$ in agreement with the DMP.

mesh	1/10	1/20	1/40	1/80
error	0.86	0.42	0.21	0.11

Table 1: H^1 -norm of the error obtained with the stabilized Galerkin approximation

To assess the convergence rate in the H^1 -norm, we consider the following problem: Ω is the unit square, and we take f and g so that the exact solution is $u(x, y) = \sin(2\pi x) \sin(2\pi y)$. The H^1 -norm of the error obtained with the scheme (4) on four consecutive meshes is presented in Table 1. The meshes are unstructured Delaunay triangulations with $h = 1/10$, $h = 1/20$, $h = 1/40$, and $h = 1/80$. Table 1 shows that the discrete solution produced by the stabilized scheme (4) is first-order in the H^1 -norm. This numerical example illustrates the fact that (4) yields approximate solutions satisfying the DMP while maintaining the same convergence order in the H^1 -norm as the standard Galerkin approximation (the convergence order in the L^2 -norm is suboptimal owing to the stabilization term). To conclude, we observe that the main purpose of the nonlinear stabilization scheme presented in this Note is not to solve Laplace-type equations since this would transform a linear problem into a nonlinear problem. Instead, it is to be used when approximating coupled systems of nonlinear PDEs; in this case, the governing equations are already nonlinear so that the stabilization term adds little complexity.

References

- [1] E. Burman and A. Ern. Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the advection-diffusion-reaction equation. *Comput. Methods Appl. Mech. Engrg.* **191**:3822–3855, 2002.
- [2] E. Burman and A. Ern. Stabilized Galerkin Approximation of Convection–Diffusion–Reaction equations: Discrete Maximum Principle and Convergence *submitted to Math. Comp.*
- [3] P.G. Ciarlet and P.-A. Raviart. Maximum principle and uniform convergence for the finite element method. *Comput. Methods Appl. Mech. Engrg.*, **2**:17–31, 1973.
- [4] J. Xu and L. Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Math. Comp.*, **66**(228):1429–1446, 1999.
- [5] S. Korotov, M. Křížek, and P. Neittaanmäki. Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle. *Math. Comp.*, **70**(233):107–119, 2000.
- [6] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics (Springer, Berlin, 2001).