On high-order approximation and stability with conservative properties

Juan Galvis[1], Eduardo Abreu[2], Ciro Díaz[2], and Marcus Sarkis[3]

1 Summary

In this paper, we explore a method for the construction of locally conservative flux fields. The flux values are obtained through the use of a Ritz formulation in which we augment the resulting linear system of the continuous Galerkin (CG) formulation in a higher-order approximation space. These methodologies have been successfully applied to multi-phase flow models with heterogeneous permeability coefficients that have high-variation and discontinuities. The increase in accuracy associated with the high order approximation of the pressure solutions is inherited by the flux fields and saturation solutions. Our formulation allows us to use the saddle point problems analysis to study approximation and stability properties as well as iterative methods design for the resulting linear system. In particular, here we show that the low-order finite element problem preconditions well the high-order conservative discrete system. We present numerical evidence to support our findings.

2 Problem and conservative formulation

Consider the equation,

$$-\mathrm{div}(\Lambda(x)\nabla p) = q \quad \text{in} \quad \Omega \subset \Re^2, \tag{1}$$

$$p = 0 \quad \text{on} \quad \partial\Omega, \tag{2}$$

[1]Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia.
[2]University of Campinas, Department of Applied Mathematics, 13.083-970, Campinas, SP, Brazil; supported by FAPESP Grant 2016/23374-1.
[3]Department of Mathematical Sciences, Worcester Polytechnic Institute Worcester USA; supported by NSF DMS-1522663

where Ω is a two-dimensional domain and Λ is a (smooth enough) positive definite symmetric matrix function. See [6] for the case of Λ being a multiscale coefficient with high-contrast. Our main interest is to obtain approximate solutions of the second order problem above[1] with: 1) high-order approximation (e.g., multiple basis per node), 2) local mass conservation properties and 3) stable-fast solver.

Our motivations come from the fact that in some applications it is **imperative** to have some conservative properties represented as conservations of total flux in control volumes. For instance, if $\mathbf{q}^h$ represents the approximation to the flux (in our case $\mathbf{q}^h = -\Lambda\nabla p^h$ where p^h is the approximation of the pressure), it is required that

$$\int_{\partial V} \mathbf{q}^h \cdot \mathbf{n} = \int_V q \quad \text{for each control volume } V.$$

For Dirichlet boundary condition, V is a control volume that does not cross $\partial\Omega$ from a set of control volumes of interest, and here and after $\mathbf{n}$ is the normal vector pointing out the control volume. We say that a discrete method is conservative if the total flux restriction such as the one written above holds.

We note that FV methods that use higher degree piecewise polynomials have been introduced in the literature; see [3, 4, 5]. We consider a Ritz formulation and construct a solution procedure that combines a continuous Galerkin-type formulation that concurrently satisfies mass conservation restrictions. We impose finite volume restrictions by using a scalar Lagrange multiplier for each restriction; see [1, 6].

The variational formulation of problem (1) is to find $p \in H_0^1(\Omega)$ such that

$$a(p, v) = F(v) \quad \text{for all } v \in H_0^1(\Omega), \tag{3}$$

where the bilinear form a is defined by

$$a(p, v) = \int_\Omega \Lambda(x)\nabla p(x)\nabla v(x)dx, \tag{4}$$

the functional F is defined by $F(v) = \int_\Omega q(x)v(x)dx$. The Problem (3) is equivalent to the minimization problem:

$$p = \arg \min_{v \in H_0^1(\Omega)} \mathcal{J}(v) \quad \text{where} \quad \mathcal{J}(v) = \frac{1}{2}a(v, v) - F(v). \tag{5}$$

Let the triangulation $\tau_h = \{R_k\}_{k=1}^{N_h}$ made of elements that are triangles or squares, where N_h is the number of elements. We also introduce the dual

[1] The use of second order formulation makes sense especially for cases where some form of high regularity holds. Usually in these cases the equality in the second order formulation is an equality in L^2 so that, in principle, there is no need to write the system of first order equations and weaken the equality by introducing less regular spaces for the pressure as it is done in mixed formulation with L^2 pressure.

mesh $\tau_h^* = \{V_k\}_{k=1}^{N_h^*}$ where the elements are called control volumes. In this paper we assume that each V_k is a subdomain of Ω with polygonal boundary. Let us introduce the space $H := \{v \in H_0^1(\Omega) : \Lambda\nabla v \in \mathrm{H}(\mathrm{div}, \Omega)\}$. If $q \in L^2$ we have that (3) is equivalent to: Find $p \in H_0^1$ such that

$$p = \arg\min_{v \in \mathcal{W}} \mathcal{J}(v), \tag{6}$$

where $\mathcal{W} = \left\{ v \in H : \int_{\partial T} -\Lambda\nabla v \cdot \mathbf{n} = \int_T q \quad \text{for all } T \in \tau_h^* \right\}$.

Problem (6) above can be view as Lagrange multipliers min-max optimization problem. See [2] and references therein. Let us denote $M_h = \mathbb{R}^{N_h^*}$.

The Lagrange multiplier formulation of problem (6) can be written as: Find $p \in H$ and $\lambda \in M_h$ that solves

$$(p, \lambda) = \arg \max_{\mu \in \mathbb{R}^{N_h^*}} \min_{v \in H,} \mathcal{J}(v) - (\overline{a}(v, \mu) - \overline{F}(\mu)). \tag{7}$$

Here, the total flux bilinear form $\overline{a} : H \times M_h \to \mathbb{R}$ is defined by

$$\overline{a}(v, \mu) = \sum_{k=1}^{N_h} \mu_k \int_{\partial V_k} \Lambda\nabla v \cdot \mathbf{n} \quad \text{for all } v \in H \text{ and } \mu \in M^h. \tag{8}$$

The functional $\overline{F} : M_h \to \mathbb{R}$ is defined by $\overline{F}(\mu) = \sum_{i=1}^{N_h} \mu_k \int_{V_k} q$, for all $\mu \in M^h$. The first order conditions of the min-max problem above give the following saddle point problem: Find $p \in H_0^1(\Omega)$ and $\lambda = 0 \in M_h$ that solves:

$$\begin{array}{ll} a(p, v) + \overline{a}(v, \lambda) = F(v) & \text{for all } v \in H, \\ \overline{a}(p, \mu) \qquad\quad = \overline{F}(\mu) & \text{for all } \mu \in M^h. \end{array} \tag{9}$$

3 Discretization and error

Let us consider $P^h = \mathbb{Q}^r(\tau_h) \cap H_0^1(\Omega)$. We also interpret M^h as $\mathbb{Q}^0(\tau_h^*)$, that is, the space of piecewise constant functions on the dual mesh τ_h^*. See for instance [6] where we consider GMsFEM spaces instead of piecewise polynomials.

The discrete version of (9) is to find $p^h \in P^h$ and $\lambda \in M^h$ such that

$$a(p^h, v^h) + \overline{a}(v^h, \lambda^h) = F(v^h) \qquad \text{for all } v^h \in P^h \tag{10}$$

$$\overline{a}(p^h, \mu^h) = \overline{F}(\mu^h) \qquad \text{for all } \mu^h \in M^h. \tag{11}$$

The equivalent matrix form is,

$$\begin{bmatrix} A & \overline{A}^T \\ \overline{A} & O \end{bmatrix} \begin{bmatrix} u^h \\ \lambda^h \end{bmatrix} = \begin{bmatrix} f \\ \overline{f} \end{bmatrix} \tag{12}$$

where A is the finite element stiffness matrix corresponding to finite element space $P^h = \text{span}\,\{\varphi_j\}$,

$$A = [a_{i,j}] \quad \text{where } a_{ij} = \int_\Omega \Lambda \nabla \varphi_i \cdot \nabla \varphi_j. \tag{13}$$

The restriction or finite volume matrix $\overline{A}$ is given by,

$$\overline{A} = [\overline{a}_{k,j}] \quad \text{where } \overline{a}_{kj} = \int_{\partial V_k} \Lambda \nabla \varphi_j \cdot \mathbf{n}. \tag{14}$$

Moreover, $f = [f_i]$ with $f_i = \int_\Omega q\,\varphi_i$ and $\overline{f} = [\overline{f}_k]_{k=1}^{N_h^*}$ with $\overline{f}_k = \int_{V_k} q$.

Note that matrix $\overline{A}$ is related to classical (low order) finite volume matrix. Matrix $\overline{A}$ is a rectangular matrix with more columns than rows. Several previous works on conservative high-order approximation of second order elliptic problem have been designed by "adding" rows using several constructions. See [1] for details.

We consider a particular case of a regular mesh made of squares. Our analysis is valid for high order finite element on regular meshes made of triangles since a similar analysis holds in this case. Define $\Gamma^* = \bigcup_{k=1}^{N_h^*}$ that is, Γ^* is the interior interface generated by the dual triangulation. For $\mu \in M^h$ define $[\mu]$ on Γ^* as the jump across element interfaces such that $[\mu]|_{\partial V_k \cap \partial V_{k'}} = \mu_k - \mu_{k'}$. Note that $\overline{a}(v, \overline{\mu}) = \sum_{k=1}^{N_h^*} \mu_k \int_{\partial V_k} \nabla v \cdot \mathbf{n} = \int_{\Gamma^*} \nabla v \cdot \mathbf{n}\,[\mu]$.

In our analysis we use the energy norm in the space that approximates the pressure and a discrete norm in the space of Lagrange multipliers. Denote $\|v\|_a^2 = \int_\Omega \Lambda \nabla v \cdot \nabla v$ for all $v \in H_0^1(\Omega)$. Let us recall the definition of space $H := \{v \in H_0^1(\Omega) : \Lambda \nabla v \in \text{H(div}, \Omega)\}$, and additional set $P_+^h = \text{Span}\{P^h, H\}$. We define the norm (that is motivated by the analysis)

$$\|v\|_{P_+^h}^2 = |v|_{H^1(\Omega)}^2 + h^2 \sum_{\ell=1}^{N_h^*} \|\Delta v\|_{L^2(R_\ell)}^2 \quad \text{for all} \quad v \in P_+^h. \tag{15}$$

Note that if $v \in \mathbb{Q}^r$, then $\|v\|_{P_+^h}^2 \asymp |v|_{H^1(\Omega)}^2$ using an inverse inequality. Also define the discrete norm for the spaces of Lagrange multipliers as

$$\|\mu\|_{M^h}^2 = \frac{1}{h} \int_{\Gamma^*} [\mu]^2. \tag{16}$$

It is possible to verify that ([1])

1. **Augmented norm**: $\|v\|_a \leq \|v\|_{P_+^h}$ for all $v \in P_+^h$.
2. **Continuity**: $|\bar{a}| \in \mathbb{R}$ such that $|\bar{a}(v,\mu^h)| \leq |\bar{a}|\|v\|_{P_+^h}\|\mu^h\|_{M^h}$ for all $v \in P_+^h$ and $\mu^h \in M^h$.
3. **Inf-Sup**: $\inf_{\mu^h \in M^h} \sup_{v \in P_+^h} \frac{\overline{a}(v,\mu^h)}{\|v\|_a \|\mu^h\|_{M^h}} \geq \alpha > 0$.

We also have established optimal approximation in energy norm ($\|p - p^h\|_a \preceq h|p|_{H^2(\Omega)}$) and using a duality argument it is possible to write the optimal L^2 approximation $\|p - (p^h + \lambda^h)\|_0 \preceq h^2|p|_{H^2(\Omega)}$; see [1] for details.

4 The case of highly anisotropic media

One issue with some cases of conservative methods is the lack of coerciveness under the presence of high-anisotropic coefficients. We can think our formulation as a stabilization for these cases (in the sense that we increase the space of the solution while keeping fixed the space for the Lagrange multipliers). Preliminary numerical studies suggest that our formulation is more robust (with respect to anisotropy) than the classical finite volume formulations.

A nice feature of our formulation is that the symmetric saddle point (12) is suitable for constructing robust preconditioners; see [2] for variety of solvers and iteration that can be used. Here we present a simple stationary iteration. Consider the iteration

$$\begin{aligned} Au_{k+1} &= f - \overline{A}^T\lambda_k \\ \lambda_{k+1} &= \lambda_k + \omega B^{-1}(\overline{A}u_{k+1} - \overline{f}). \end{aligned} \tag{17}$$

Here ω is a relaxation parameter and B a preconditioner to be defined. This iteration corresponds to a preconditioned Richardson iteration applied to the Schur complement problem (to solve for the Lagrange multiplier lambda equation). We have, by combining the two equations above,

$$\lambda_{k+1} = \lambda_k + \omega B^{-1}\left(g - S\lambda_k\right)$$

where $g = \overline{A}A^{-1}f - \overline{f}$ and S is the Schur complement $S = \overline{A}A^{-1}\overline{A}^T$. Note that the size of S is the number of interior vertices if the control volumes are constructed by joining the centers of the elements of the primal mesh. In the case of isotropic coefficients and square elements, we can take $B = M_h$ defined in (16); see [2]. In order to take into account the anisotropy, below in the numerical tests we consider B defined by

$$B = [b_{ij}] \text{ where } b_{ij} = \int_D \Lambda\nabla\varphi_i\nabla\varphi_j \text{ with } \varphi_i, \varphi_j \in \mathbb{Q}^1 \cap H_0^1(\Omega).$$

5 Numerical experiments

We consider the Dirichlet problem (1). Let $\Omega = (0,1) \times (0,1)$. We consider a regular mesh made of 4^L squares. The dual mesh is constructed by joining the centers of the elements of the primal mesh. We perform a series of numerical experiments to compare properties of FEM solutions with the solution of our high order FV formulation (to which we refer from now on as FV solution). We select the exact solution $p(x,y) = \sin(\pi x)\sin(\pi y)(-x+3y)$ and $f = -\Delta u$.

On Table 1 we compare our $\mathbb{Q}^1$ FV method with the classical $\mathbb{Q}^1$ finite element method. We compute L^2 and H^1 errors. We observe optimal convergence of both strategies however the FV is conservative. On Table 2 we consider $\mathbb{Q}^2$ elements and optimal higher convergence rates are confirmed.

L	$FEM,\ L^2\ Error$	$FV.\ L^2\ Error$	$FEM,\ H^1\ Error$	$FV.\ H^1\ Error$
1	1.5538×10^{-1}	1.5103×10^{-1}	1.1297×10^{0}	1.1338×10^{0}
2	3.6342×10^{-2}	3.1881×10^{-2}	5.3226×10^{-1}	5.3416×10^{-1}
3	8.9720×10^{-3}	$7.5.276 \times 10^{-3}$	2.6374×10^{-1}	2.6403×10^{-1}
4	2.2548×10^{-3}	1.9348×10^{-3}	1.3163×10^{-1}	1.3172×10^{-1}
5	5.5513×10^{-4}	4.6095×10^{-4}	6.5833×10^{-2}	6.5840×10^{-2}
6	1.3875×10^{-4}	1.1513×10^{-4}	3.2948×10^{-2}	3.2924×10^{-2}
7	3.4685×10^{-5}	2.8776×10^{-5}	1.6418×10^{-2}	1.6489×10^{-2}

Table 1 Table of **FEM** and **FV** L^2 and H^1 errors using $\mathbb{Q}^1$ elements.

L	$FEM\ L^2\ Error$	$FV.\ L^2\ Error$	$FEM\ H^1\ Error$	$FV.\ H^1\ Error$
1	1.4061×10^{-2}	2.4548×10^{-2}	1.9302×10^{-1}	2.2436×10^{-1}
2	2.1217×10^{-3}	4.9023×10^{-3}	5.4862×10^{-2}	7.2895×10^{-2}
3	2.6860×10^{-4}	6.4789×10^{-4}	1.4072×10^{-2}	1.8847×10^{-2}
4	3.3875×10^{-5}	8.1756×10^{-5}	3.5418×10^{-3}	4.7552×10^{-3}
5	4.2437×10^{-6}	1.0242×10^{-5}	8.3539×10^{-4}	1.2667×10^{-3}
6	5.3075×10^{-7}	1.2810×10^{-6}	2.2016×10^{-4}	2.9616×10^{-4}
7	6.6353×10^{-8}	1.6015×10^{-7}	5.5043×10^{-5}	7.4046×10^{-5}

Table 2 Table of **FEM** and **FV** L^2 and H^1 errors using $\mathbb{Q}^2$ elements.

We now move to symmetric anisotropic coefficients Λ. We now show in Tables 3-8 the smallest and the largest eigenvalues of $\lambda\max(B^{-1}S)/\lambda_{\min}(B^{-1}S)$ for different values of Λ, $h = 2^L$ and for $\mathbb{Q}^1, \mathbb{Q}^2$ and $\mathbb{Q}^3$ elements. The Λ has eigenvalues 1 and η and associate eigenvector $\eta = (\cos(\Theta), \sin(\Theta))^t$. From these results we see that the smallest eigenvalue is very stable, therefore, the discrete inf-sup is satisfied. This is a strong result since finite volume discretizations sometimes lack in coerciveness for highly anisotropic media. The proposed preconditioner performs well however has a mildly dependence with respect to the different configuration of anisotropy direction and anisotropy

ratio. This is somehow expected since the continuity given in (15) is with respect to the V_h-norm rather than a-norm, and further studies are on the way to eliminate this dependence. Recall that the application of the preconditioner requires the solution of a low-order ($\mathbb{Q}^1$) classical symmetric finite element problem. In practice, these solve can be replaced by a robust method for low-order finite element method and inexact Uzawa or Conjugated Gradient. Recall also that we obtain conservative solutions.

$L\backslash\eta$	1	10	100	1000	1	10	100	1000	1	10	100	1000
2	1.76	1.76	1.76	1.76	1.76	1.76	1.81	1.81	1.76	1.80	1.82	1.83
	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
3	2.09	2.09	2.09	2.09	2.09	2.11	2.12	2.12	2.09	2.11	2.13	2.14
	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
4	2.20	2.20	2.20	2.20	2.20	2.21	2.22	2.22	2.20	2.21	2.22	2.22
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	2.24	2.24	2.24	2.24	2.23	2.24	2.24	2.24	2.24	2.24	2.24	2.24
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3 Maximum and minimum eigenvalue $\frac{\lambda_{max}}{\lambda_{min}}$ for $\Theta = 1$(left), $\Theta = \frac{\pi}{6}$ (center), $\Theta = \frac{\pi}{4}$ (right) and $P^h = \mathbb{Q}^1$. The Λ has eigenvalues 1 and η and the eigenvector associated to η is $(\cos(\Theta), \sin(\Theta))^t$.

$L\backslash\eta$	1	10	100	1000	1	10	100	1000	1	10	100	1000
2	1.79	1.80	1.81	1.81	1.79	2.13	2.47	2.53	1.79	2.32	2.98	3.12
	1.05	1.05	1.06	1.06	1.05	1.09	1.11	1.11	1.05	1.10	1.12	1.12
3	2.10	2.10	2.11	2.11	2.10	2.50	2.99	3.18	2.10	2.77	4.03	4, 43
	1.01	1.01	1.01	1.01	1.01	1.02	1.03	1.03	1.01	1.02	1.03	1.03
4	2.21	2.21	2.21	2.21	2.21	2.61	3.27	3.92	2.21	2.91	4.40	5.24
	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.00	1.01	1.01	1.01
5	2.24	2.24	2.24	2.24	2.24	2.64	3.43	4.90	2.24	2.95	4.52	6.43
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	2.25	2.25	2.25	2.25	2.25	2.65	3.48	5.86	2.25	2.95	4.57	7.60
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4 Maximum and minimum eigenvalue $\frac{\lambda_{max}}{\lambda_{min}}$ for $\Theta = 1$ (left), $\Theta = \frac{\pi}{6}$ (center), $\Theta = \frac{\pi}{4}$ (right) and $P_h = \mathbb{Q}^2$. The Λ has eigenvalues 1 and η and the eigenvector associated to η is $(\cos(\Theta), \sin(\Theta))^t$.

6 Conclusions

In this paper we use a Ritz formulation with constraints to obtain locally conservative fluxes in the approximation of the Darcy equation. With this

$L\backslash\eta$	1	10	100	1000	1	10	100	1000	1	10	100	1000
2	4.52	4.54	4.57	4.58	4.52	4.72	6.92	7.21	4.52	4.80	5.01	5.05
	2.43	2.43	2.43	2.43	2.29	2.29	2.29	2.29	2.44	2.28	2.24	2.23
3	5.32	5.35	5.40	5.41	5.33	5.47	6.92	7.21	5.32	5.67	7.37	7.68
	2.29	2.29	2.29	2.29	2.29	1.95	1.89	1.90	2.29	1.92	1.86	1.64
4	5.59	5.62	5.68	5.69	5.60	5.89	10.8	12.2	5.59	6.48	12.2	13.7
	2.26	2.26	2.26	2.26	2.26	1.81	1.74	1.74	2.26	1.78	1.72	1.72
5	5.67	6.70	5.75	5.77	5.67	6.19	16.9	22.2	5.67	7.09	20.2	26.3
	2.25	2.25	2.25	2.25	2.25	1.75	1.68	1.67	2.25	1.73	1.67	1.66
6	5.69	5.72	5.77	5.79	5.68	5.68	24.7	41.4	5.68	7.46	30.8	50.7
	2.25	2.25	2.25	2.25	2.25	1.72	1.65	1.65	2.25	1.70	1.65	1.64

Table 5 Maximum and minimum eigenvalue $\frac{\lambda_{max}}{\lambda_{min}}$ for $\Theta = 1$ (left), $\Theta = \frac{\pi}{6}$ (center), $\Theta = \frac{\pi}{4}$ (right) and $P_h = \mathbb{Q}^3$.

formulation we obtain solution that have high-order approximation and still yield locally conservative fluxes with no post-processing. We show that the resulting linear system can be solve using a stationary iteration where the application of the preconditioner uses an approximation of a low-order finite element problem. We present numerical evidence to support our findings.

Acknowledgments E. Abreu thanks financial support FAPESP through grant No. 2016/23374-1.

References

[1] E. Abreu, C. Díaz, J. Galvis and M. Sarkis, On high-order conservative finite element methods, Computers & Mathematics with Applications. Online December 2017 (https://doi.org/10.1016/j.camwa.2017.10.020).
[2] M. Benzi, G. H. Golub and J. Liesen, Numerical solution of saddle point problems, Acta numerica 14 (2005) 1-137.
[3] L. Chen, A New Class of High Order Finite Volume Methods for Second Order Elliptic Equations, SIAM J. Numer. Anal., 47(6) (2010) 4021-4043.
[4] Z. Chen, J. Wu and Y. Xu, Higher-order finite volume methods for elliptic boundary value problems, 37(2) (2012) 191-253.
[5] Z. Chen, Y. Xu and Y. Zhang, A construction of higher-order finite volume methods, Math. Comp. 84 (2015) 599-628.
[6] M. Presho and J. Galvis, A mass conservative Generalized Multiscale Finite Element Method applied to two-phase flow in heterogeneous porous media, Journal of Computational and Applied Mathematics, 296 (2016) 376-388.