A multigrid method for mixed finite element discretizations of current continuity equations

Citation for published version (APA):

Document status and date:
Published: 01/01/1990

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
A MULTIGRID METHOD FOR
MIXED FINITE ELEMENT
DISCRETIZATIONS OF CURRENT
CONTINUITY EQUATIONS

by

A. Reusken

Reports on Applied and Numerical Analysis
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
Abstract. Mixed finite element discretizations for current continuity equations are presented in [6, 8, 12]. We consider the resulting system of equations and develop a multigrid method for such a system.

1. Introduction

Recently (new) mixed finite element schemes for current continuity equations have been presented in [6, 8, 12]. These schemes have nice properties: they provide an $M$-matrix, there is current preservation, and a good approximation of sharp shapes. Such a scheme results in a large sparse system of equations for the unknowns. A "standard" multigrid solver cannot be used for this system due to the presence of (extremely) large convection in part of the domain and the use of mixed finite elements. Our goal here is to develop a suitable multigrid method for this system.

First we investigate a relation between the mixed finite elements used and nonconforming Crouzeix-Raviart finite elements. Using this relation and the multigrid theory for nonconforming finite elements in [3], [5] then leads to a multigrid method for the discretized current continuity equations.

This paper only deals with the derivation of the method. In a forthcoming paper numerical results will be presented.

The following system of scaled equations is often used for semiconductor device simulation in the case of a stationary problem with constant coefficients (cf. [9], [10]):

\[
\begin{align*}
\Delta \psi &= n - p + D & \text{in } \Omega \subset \mathbb{R}^2 & \quad (a) \\
\text{div}(\nabla n - n \nabla \psi) &= R & \text{in } \Omega & \quad (b) \\
\text{div}(\nabla p + p \nabla \psi) &= R & \text{in } \Omega & \quad (c) \\
+ \text{boundary conditions}.
\end{align*}
\]

The unknowns $\psi, n$ and $p$ represent the (scaled) electric potential and the (scaled) concentration of negative and positive charges respectively. $D$ describes the doping profile and $R$ the generation-recombination term. The domain $\Omega$ is scaled to have a diameter of 1.

We assume that for solving (1) the equations are decoupled in some iterative method.
and we restrict ourselves to the equation (b) or (c) with a given potential $\psi$. Moreover we only consider the situation in which (a linearization of) $R$ in equation (b) (or (c)) is independent of $n$ (or $p$). It is possible to weaken this restriction by using the theory in [8].

In the remainder we consider the following problem for the concentration of positive charges $u$:

\[
\begin{align*}
\text{Find } u & \in H^1(\Omega) \text{ such that} \\
\text{div}(\nabla u + u\nabla \psi) &= f \quad \text{in } \Omega \subset \mathbb{R}^2 \\
u &= g \quad \text{on } \Gamma_0 \subset \partial \Omega \\
\frac{\partial u}{\partial n} + u \frac{\partial \psi}{\partial n} &= 0 \quad \text{on } \Gamma_1 = \partial \Omega \setminus \Gamma_0.
\end{align*}
\]

In this current continuity equation (the current is defined by $J = \nabla u + u\nabla \psi$) we assume that $\psi$ is a given function.

2. Current continuity equation and mixed finite elements

In this section we give the discretization of the current continuity equation (2) as presented in [6].

Rewriting (2) in terms of the Slotboom variable $\rho := e^\psi u$ results in the following problem:

\[
\begin{align*}
\text{Find } \rho & \in H^1(\Omega) \text{ such that} \\
\text{div}(e^{-\psi} \nabla \rho) &= f \quad \text{in } \Omega \\
\rho &= \chi := e^\psi g \quad \text{on } \Gamma_0 \\
\frac{\partial \rho}{\partial n} &= 0 \quad \text{on } \Gamma_1.
\end{align*}
\]

For ease we assume that $\Omega$ is a polygonal domain. Let $\{T_k\}_{k \geq 0}$ be a regular sequence of decompositions of $\Omega$ into triangles $T$.

Now use the following well-known lowest order Raviart-Thomas finite element spaces. Set $RT(T) := \{\tau = (\tau_1, \tau_2) \mid \tau_1 = \alpha + \beta x, \tau_2 = \gamma + \beta y, \alpha, \beta, \gamma \in \mathbb{R} \} \quad (T \in T_k)$, and define $V_k := \{\tau \in (L^2(\Omega))^2 \mid \text{div} \tau \in L^2(\Omega), \tau \cdot n = 0 \text{ on } \Gamma_1, \tau|_T \in RT(T) \text{ for all } T \in T_k\}$ and also $W_k := \{\phi \in L^2(\Omega) \mid \phi|_T \in P_0(T) \text{ for all } T \in T_k\}$.

A mixed finite element discretization of (3) is as follows:

\[
\begin{align*}
\text{Find } \mathcal{J}_k & \in V_k \text{ and } \rho_k \in W_k \text{ such that} : \\
\int_{\Omega} e^\psi \mathcal{J}_k \cdot \tau \, dx + \int_{\Gamma_0} \rho_k \text{div} \tau \, d\Gamma &= \int_{\Gamma_1} \chi \tau \cdot n \, d\Gamma \quad \forall \tau \in V_k \\
\int_{\Omega} \phi \text{div} \mathcal{J}_k \, dx &= \int_{\Omega} f \phi \, dx \quad \forall \phi \in W_k.
\end{align*}
\]
This problem has a unique solution and every $\tau \in V_k$ has a continuous normal component when passing from one element to another, so in particular the current $J_k$ is preserved. In the remainder we assume that $\psi = \psi_k$ is given and is continuous on $\Omega$ and linear on every $T \in T_k$.

The system resulting from (4) has the form \[
\begin{pmatrix}
A & B \\
B^T & 0
\end{pmatrix}
\] and is not positive definite. One might try to solve this system by using a solver that is adapted to saddle-point problems (see e.g. [2], [4]). However in the situation here one then expects serious trouble due to the large differences in scaling of $A$ and $B$ and it is not clear how to overcome these scaling problems. Further investigations in this direction have not been done (yet). Another possibility (used in [6]) is to introduce Lagrange-multipliers and use static condensation resulting in a system (for the Lagrange-multipliers) with a symmetric positive definite $M$-matrix. For this problem a suitable rescaling can be used and the final system then still has an $M$-matrix.

The latter approach is what we use.

Let $E_k$ be the set of edges of $T_k$.

Define $\tilde{V}_k := \{ \tau \in \{ L^2(\Omega) \}^2 \mid \tau|_T \in RT(T) \mbox{ for all } T \in T_k \}$, and for $\xi \in L^2(\Gamma_0)$

\[
\Lambda_{k,\xi} := \{ \mu \in L^2(E_k) \mid \mu|_e \in P_0 \mbox{ for all } e \in E_k, \int \mu(\gamma) ds = 0 \mbox{ for all } e \in E_k \mbox{ with } e \subset \Gamma_0 \}.
\]

Now consider the following problem:

\[
\begin{align*}
\text{Find } & \tilde{J}_k \in \tilde{V}_k, \quad \tilde{\rho}_k \in W_k, \quad \tilde{\lambda}_k \in \Lambda_{k,\xi} \text{ such that } \\
& \int e^{\gamma} \tilde{J}_k \cdot \tau \, dx + \sum T \int \tilde{\rho}_k \, \text{div} \tau \, dx - \sum T \int \tilde{\lambda}_k \tau \cdot n \, ds = 0 \quad \forall \tau \in \tilde{V}_k \\
& \sum T \int \phi \, \text{div} \tilde{J}_k \, dx = \int f \phi \, dx \quad \forall \phi \in W_k \\
& \sum T \int \mu \tilde{J}_k \cdot n \, ds = 0 \quad \forall \mu \in \Lambda_{k,\xi}.
\end{align*}
\] (From the context it is clear that summation should be taken over all $T \in T_k$.)

This problem has a unique solution and $\tilde{J}_k \equiv J_k, \quad \tilde{\rho}_k \equiv \rho_k$ holds. Moreover, $\tilde{\lambda}_k$ is a good approximation of $\rho$ at the interelements (see [1]). In the resulting matrix-vector problem the unknowns corresponding to $\tilde{J}_k$ and $\tilde{\rho}_k$ can be eliminated element by element, by static condensation. The resulting problem for $\tilde{\lambda}_k$ can also be derived from the variational formulation in (5). This is done in the lemma below.

For $g \in L^2(\tilde{T})$ with $\tilde{T} \supset T$ we use the notation $\tilde{g}|_T := \frac{1}{|T|} \int_{\tilde{T}} g(z) \, dz$.

Lemma 2.1. For $f$ as in (5) (or (2)) define $\tilde{I}_f$ by $\tilde{I}_f'(\tilde{z}) = \frac{1}{2} \tilde{f}|_{\tilde{T} \supset T}$. Define a symmetric bilinear form $b_k(\cdot, \cdot)$ and a linear functional $g_k(\cdot)$ on $L^2(E_k)$ as follows:

\[
b_k(\lambda, \mu) = \sum T (|T|e^{\gamma}|_T)^{-1} \int \lambda n \, ds \cdot \int \mu n \, ds
\]
The solution \( \tilde{\lambda}_k \) of (5) is also the unique solution of the following problem:

\[(6A) \quad \begin{cases} 
\text{Find } \tilde{\lambda}_k \in \Lambda_{k,C} \text{ such that } \\
b_k(\tilde{\lambda}_k, \mu) = g_k(\mu) \text{ for all } \mu \in \Lambda_{k,0}.
\end{cases}
\]

**Proof.** First note that (6A) has a unique solution because \( b_k(\cdot, \cdot) \) is positive definite on \( \Lambda_{k,0} \). Now consider (5) with solution \( \tilde{f}_k, \tilde{\rho}_k, \tilde{\lambda}_k \). The definition of \( \tilde{V}_k \) implies that \( \tilde{f}_k = (\alpha, \gamma) + \beta(x_1, x_2) =: \zeta_T + \beta \hat{z} \).

So \( \text{div} \tilde{f}_k = 2\beta \). The second equation in (5) implies that \( \int \beta \, dx = \int \beta \, dx \), so \( \beta = \frac{1}{2} \) and \( \tilde{f}_k = \zeta_T + \beta \hat{f} \). By taking \( \zeta_T = (0,1) \) and \( \beta \hat{f}_T = (1,0) \) (and zero outside \( T \)) the first equation in (5) yields

\[
\zeta_T \int_T e^\psi \, dx = \int_{\partial T} \tilde{\lambda}_k \hat{n} \, ds - \int_T e^\psi \, J \hat{f} \, dx,
\]

so

\[
\zeta_T = (|T|e^\psi) \int_T \tilde{\lambda}_k \hat{n} \, ds - \int_T e^\psi \, J \hat{f} \, ds.
\]

Substituting \( \tilde{f}_k = \zeta_T + \beta \hat{f} \) in the third equation of (5) we see that \( \tilde{\lambda}_k \) satisfies \( b_k(\tilde{\lambda}_k, \mu) = g_k(\mu) \) for all \( \mu \in \Lambda_{k,0} \). \( \square \)

For formulating the problem (6A) in \( \mathbb{R}^n \) we take the standard basis \( \{ \mu_i \}_{i \in I \cup I_0} \) in \( \Lambda_k = \{ \mu \in L^2(E_k) \mid \mu|_e \in P_0 \text{ for all } e \in E_k \} \) with \( I_0 \) the index set corresponding to edges \( e_i \subset \Gamma_0 \) and \( I \) the index set of edges \( e_i \subset \Omega \setminus \Gamma_0 \). We write \( \lambda_k = \lambda_k^l + \lambda_k^d \) with \( \lambda_k^l = \sum_{i \in I_0} \bar{\chi}_{|e_i} \mu_i \) and \( \lambda_k^d = \sum_{i \in I} \alpha_i \mu_i \). This then results in the following systems of equations for the unknowns \( \{ \alpha_i \}_{i \in I} \):

\[(6B) \quad \sum_{j \in I} b_k(\mu_j, \mu_i) \alpha_j = - \sum_{j \in I_0} b_k(\mu_j, \mu_i) \bar{\chi}_{|e_j} + g_k(\mu_i) \quad \text{for} \quad i \in I.
\]

**Remark 2.2.** For the stiffness matrix \( A \) corresponding to (6B) we get

\[
A_{ij} = b_k(\mu_j, \mu_i) = \sum_T (|T|e^\psi) \int_T \mu_j \hat{n} \, ds \cdot \int_T \mu_i \hat{n} \, ds
= \sum_T (e^\psi) \int_T (|e_j|n_T^{(j)} \cdot (|e_i|n_T^{(i)})) \text{ where } n_T^{(m)} \text{ is nonzero only if } e_m \subset \partial T \text{ in which case it equals the outward unit normal corresponding to triangle } T \text{ and edge } e_m.
\]
3. Properties of the stiffness matrix

In this section we prove some important properties of the stiffness matrix corresponding to (6B). We start with a lemma about elementary geometric properties of a triangle.

**Lemma 3.1.** Consider the geometry as in fig. 1 below (angles might be obtuse too).

![Fig. 1.](image)

Let \( \mu(i) := |e_i| \mu(i) \) (\( i = 1, 2, 3 \)).

The following holds (with \( i, j, k \in \{1, 2, 3\}; \overline{AB} = \overline{BA} = \overline{B - A} \)):

(a) if all angles \( \leq \frac{\pi}{2} \) then \( \mu(i) \cdot \mu(j) \leq 0 \) for all \( i \neq j \)

(b) \( |T| = \mu(i) \cdot m_j m_k \) for \( i \neq j \)

(c) \( \mu(1) + \mu(2) + \mu(3) = 0 \)

(d) Let \( \phi_i \) be linear on \( T \), \( \phi_i(m_i) = 1 \), \( \phi_i(m_r) = 0 \) for all \( r \in \{1, 2, 3\} \setminus \{i\} \), then \( \phi_i(T)(z) = |T|^{-1} \mu(i) \cdot (z - m_r) \) for \( r \in \{1, 2, 3\} \setminus \{i\} \)

(e) Let \( \phi \in P_2(T) \), then \( \int_T \phi(z)dz = \frac{|T|}{3}(\phi(m_1) + \phi(m_2) + \phi(m_3)) \).

**Proof.** (a) is obvious. Concerning (b), it is no restriction to take \( i = 1, j = 2 \). Now \( |T| = \frac{1}{2} \det(Q_1Q_2Q_3Q_2) = \det(m_2 m_1 m_3) \cdot |Q_3Q_2Q_1(\mu(1))| = \pm m_2 m_1 \cdot \mu(1) \); the "+" sign holds because \( m_2 m_1 \cdot \mu(1) > 0 \). So (b) holds.

Clearly \( Q_1Q_3 + Q_3Q_2 + Q_2Q_1 = 0 \) holds; now rotating over \( \frac{\pi}{3} \) we get \( \mu(2) + \mu(3) + \mu(1) = 0 \). So (c) holds. With respect to (d) note that \( \phi_i(z) = |T|^{-1} \mu(i) \cdot (z - m_r) \) is linear on \( T \); furthermore it is easy to check (using (b)) that \( \phi_i(m_r) = 1 \) and \( \phi_i(m_r) = 0 \) if \( r \neq i \).

It is easy to verify that (e) holds in the situation where \( T \) equals the reference triangle \( \hat{T} \) with vertices \((0,0), (1,0), (0,1)\). Using the affine transformation \( F : \hat{T} \rightarrow T \), \( F(z) = Bz + Q_{11} \) with \( B = (Q_1Q_2 Q_1Q_3) \) we get \( \int_T \phi(z)dz = |\det(B)| \int_T f(\phi \circ F)(y)dy = 2|T| \frac{1}{2}((\phi \circ F)(\frac{1}{2}, \frac{1}{2}) + (\phi \circ F)(0, \frac{1}{2}) + (\phi \circ F)(\frac{1}{2}, 0)) = \frac{|T|}{3}(\phi(m_1) + \phi(m_2) + \phi(m_3)) \).

\[ \Box \]

**Theorem 3.2.** The stiffness matrix corresponding to (6B) is symmetric positive definite. If the triangulation is weakly acute (all angles \( \leq \frac{\pi}{2} \)) then this matrix is an \( M \)-matrix.
Proof The bilinear form $b_k(\cdot, \cdot)$ is symmetric positive definite on $Ak_0$, so the stiffness matrix is symmetric positive definite.

We use the following criterion to prove that the stiffness matrix $A$ is an $M$-matrix (cf. [11]): if $A$ is such that

(a) $A_{ii} > 0$ for all $i$ and $A_{ij} \leq 0$ for all $i, j$ with $i \neq j$

(b) $A$ is irreducibly diagonally dominant,

then $A$ is an $M$-matrix.

It is easy to verify (a) using $A_{ij} = \sum_T (\int e^\phi dx)^{-1} e^{i(i)} \cdot e^{j(j)}$ (cf. Remark 2.2 and Lemma 3.1). Now take a triangle $T$ with $\partial T = e_i \cup e_j \cup e_k$. It is easy to verify that if $A_{ij} = 0$ then $A_{ik}A_{jk} \neq 0$, so the three unknowns related to the midpoints $m_i, m_j, m_k$ are connected by a path of nonzero coefficients. From this it follows that $A$ is irreducible. Consider the element stiffness matrix $A(T)$, i.e. $A_{ij}^{(T)} = (\int e^\phi dx)^{-1} e^{i(i)} \cdot e^{j(j)}$. If $\partial T \cap \Gamma_0 = \emptyset$ then using Lemma 3.1(c) it follows that $e^{i(i)} \cdot e^{j(j)} = -(e^{i(i)} \cdot e^{j(j)} + e^{k(k)} \cdot e^{j(j)})$ so $|A_{ii}^{(T)}| = |A_{ij}^{(T)}| + |A_{ik}^{(T)}| = \sum_{m \neq i} |A_{im}^{(T)}|$. One can also check that for a suitable $i$ with $e_i \cap \Gamma_0 = \emptyset$ and $e_j \subset \Gamma_0$ (so $\partial T \cap \Gamma_0 \neq \emptyset$) we get $|A_{ii}^{(T)}| > |A_{ik}^{(T)}| = \sum_{m \neq i} |A_{im}^{(T)}|$. Obviously these inequalities concerning diagonal and off-diagonal coefficients of the stiffness matrices $A(T)$ also hold for $A$. From this, and because $A$ is irreducible we conclude (b).

4. Another interpretation of the system given in (6B)

In this section we show that using a modified nonconforming linear finite element discretization for the problem (3) results in a system of equations that is equal to the system in (6B).

We consider the Crouzeix-Raviart (P1) element corresponding to $T_k$:

$S_k := \{ v \in L^2(\Omega) \mid v_T \text{ is linear for all } T \in T_k, v \text{ is continuous at the midpoints of edges} \}$.

As in section 2 we use an indexset $I \cup I_0$ for the edges $e_i \in E_k$. The midpoint on an edge $e_i$ is denoted by $m_i$ ($i \in I_0 \iff m_i \in I_0$ holds). The standard basis of $S_k$ is denoted by $\{ \phi_i \}_{i \in I(k)}$. For $\xi \in L^2(\Gamma_0)$ we define $S_{k, \xi} := \{ v \in S_k \mid v(m_i) = \xi_{m_i} \text{ if } i \in I_0 \}$.

Theorem 4.1. Let $f$ be as in (2) and define $G_k(f) \in L^2(\Omega)$ by $G_k(f)|_T = \int_T (1 + \frac{1}{2}(e^\psi|_T)^{-1} e^\psi).$ Consider the following problem:

\[
(7A) \begin{cases} 
\text{Find } \tilde{\eta}_k \in S_{k, \xi} \text{ such that} \\
\sum_T \int_T (e^\psi|_T)^{-1} \nabla \tilde{\eta}_k \cdot \nabla \phi \, dx = -\int_{\Omega} G_k(f) \phi \, dx \quad \text{for all } \phi \in S_{k, 0}.
\end{cases}
\]
Write \( \tilde{\eta}_k = \eta_k^d + \eta_k^f \) with \( \eta_k^d = \sum_{j \in I_k} \tilde{x}(\phi_j) \) and \( \eta_k^f = \sum_{j \in I} \alpha_k \phi_j \).

Taking \( \phi = \phi_i \) (\( i \in I \)) in \( (7A) \) and using \( \tilde{\eta}_k = \eta_k^d + \eta_k^f \) results in a system of equations for the unknowns \( \{ \alpha_j \}_{j \in I} \) that is equal to the system \( (6B) \).

**Proof.** Define the bilinear form \( a_k(\eta, \phi) = \sum_{T \in T} \int_T (\tilde{\eta}/|T|)^{-1} \nabla \eta \cdot \nabla \phi \, dx \quad (\eta, \phi \in S_k) \).

Substituting \( \tilde{\eta}_k = \eta_k^d + \eta_k^f \) in \( (7A) \) and taking \( \phi = \phi_i \) results in:

\[
(7B) \quad \sum_{j \in I} a_k(\phi_j, \phi_i) \alpha_j = - \sum_{j \in I_k} \tilde{a}_k(\phi_j, \phi_i) \tilde{x}(\phi_j) - \int_{\tilde{\eta}} G_k(f_\phi) \, dx \quad \text{for } i \in I.
\]

Comparing \( (6B) \) and \( (7B) \) we see that we only have to show:

1. \( \tilde{a}_k(\phi_j, \phi_i) = b_k(\mu_j, \mu_i) \) and
2. \( g_k(\mu_i) = - \int_{\tilde{\eta}} G_k(f_\phi) \, dx \).

Using the definitions and checking per triangle we see that it is sufficient to prove:

1. \( (a') \quad |T|^2 \left( \nabla \phi_j \right)_T \cdot \left( \nabla \phi_i \right)_T = \int_{\partial T} \mu_j n \, ds \cdot \int_{\partial T} \mu_i n \, ds \quad \text{and} \)
2. \( (b') \quad (|T|e^{\phi}_T)^{-1} \int_T e^{\phi} I^f \, dx \cdot \int_{\partial T} \mu_i n \, ds - \int_{\partial T} \mu_i I^f \cdot n \, ds = - \int_{\tilde{\eta}} G_k(f_\phi) \, dx \).

The righthand side of \( (a') \) equals \( \nu^{(j)} \cdot \nu^{(i)} \) (see Lemma 3.1 for notation).

Using Lemma 3.1(d) we see that

\[
|T|^2 \left( \nabla \phi_j \right)_T \cdot \left( \nabla \phi_i \right)_T = |T|^2 (|T|^{-1} \nu^{(j)}) \cdot (|T|^{-1} \nu^{(i)}) = \nu^{(j)} \cdot \nu^{(i)}.
\]

So \( (a') \) holds. The proof of \( (b') \) runs as follows.

\[
\frac{1}{2} \int_T \left( |T|^{-1} e^{\phi}_T \right)^{-1} e^{\phi} I^f \, dx \cdot \int_{\partial T} \mu_i n \, ds - \int_{\partial T} \mu_i I^f \cdot n \, ds
\]

\[
= \frac{1}{2} \int_T \left( |T|^{-1} e^{\phi}_T \right)^{-1} e^{\phi} (I^f \cdot \nu^{(i)}) \, dx - \frac{1}{|e_i|} \int_{\partial e_i} \nu \cdot \nu^{(i)} \, ds
\]

\[
= \frac{1}{2} \int_T \left( |T|^{-1} e^{\phi}_T \right)^{-1} e^{\phi} (|T| \phi_i + \nu^{(i)} \cdot m_j) \, dx - \frac{1}{|e_i|} \int_{\partial e_i} |T| \phi_i + \nu^{(i)} \cdot m_j \, ds
\]

(\text{use Lemma 3.1 (d)})

\[
= \frac{1}{2} \int_T \left( e^{\phi}_T \right)^{-1} e^{\phi} \phi_i \, dx + \nu^{(i)} \cdot m_j - |T| - \nu^{(i)} \cdot m_j \quad \text{(use } \phi_i \equiv 1 \text{ on } e_i)\]
\[ \frac{1}{2} \int_{T} \{ \int (e^{\psi}|T|)^{-1} e^{\psi} \phi_i \, dx - 3 \int \phi_i \, dx \} \quad \text{(use Lemma 3.1 (e))} \]

\[ = - \int_{T} f_j (1 \frac{1}{2} - \frac{1}{2} (e^{\psi}|T|)^{-1} e^{\psi}) \phi_i \, dx \]

\[ = - \int_{T} G_k(f) \phi_i \, dx \] \hspace{1cm} \square

**Remark 4.2.** From Theorem 4.1 we conclude that the system (6B) for the Lagrange multiplier \( \lambda_h \) (which is an approximation of the solution \( \rho \) of (3)) can also be obtained as the result of a modified linear nonconforming finite element discretization of (3). The modification consists of replacing \( e^{-\psi}|T| \) by \((e^{\psi}|T|)^{-1}\) and \( f|T| \) by \( f_j (1 \frac{1}{2} - \frac{1}{2} (e^{\psi}|T|)^{-1} e^{\psi}) \). Similar relations between mixed finite element discretizations and nonconforming finite element discretizations are discussed in [1].

---

### 5. A multigrid method for nonconforming finite element discretizations

This section is based on [3] and [5]. A multigrid method for solving second order elliptic boundary value problems using Crouzeix-Raviart \((P1)\) nonconforming finite elements is developed there. In particular, special restriction and prolongation operators are constructed. As in [3] we restrict ourselves to the Poisson equation. In view of the preceding sections we take mixed boundary conditions here.

We use the notation \( (u, v) := \int \nabla u \cdot \nabla v \, dx \) \( (u, v) \in H^1(\Omega) \) and \( a(u, v) := \int \nabla u \cdot \nabla v \, dx \) \( (u, v) \in H^1(\Omega) \). Let \( H^1_0(\Omega) := \{ v \in H^1(\Omega) | v = \gamma \text{ on } \Gamma_0 \} \), and \( S_k, S_k, \} \) as in §4 with \( \{ \phi_i \}_{i=1}^{n_0} \) the standard basis of \( S_k \). For \( u, v \in S_k \) we define \( a_k(u, v) := \sum_{T} \int_T \nabla u \cdot \nabla v \, dx \).

Now the following problem (with \( g \in L^2(\Omega) \) and \( \gamma \in L^2(\Gamma_0) \) given):

\[ \begin{cases} 
\text{Find } u \in H^1_0(\Omega) \text{ such that} \\
\quad a(u, \phi) = (g, \phi) \quad \text{for all } \phi \in H^1_0(\Omega) 
\end{cases} \]

\[ \text{(8)} \]

can be discretized using Crouzeix-Raviart elements, as follows:

\[ \begin{cases} 
\text{Find } u_h^* \in S_{k,0} \text{ such that} \\
\quad a_k(u_h^*, \phi) = -a_k(u_h^d, \phi) + (g, \phi) =: g_h(\phi) \quad \text{for all } \phi \in S_{k,0},
\end{cases} \]

\[ \text{(9)} \]

with \( u_h^d := \sum_{j \in I_0} \bar{\gamma}_{c_j} \phi_j \).

8
Remark 5.1. Up to now we assumed that the endpoints of \( \Gamma_0 \) coincide with vertices in the triangulation \( T_k \). However, in the multigrid method that we study below we use (very) coarse triangulations too, for which this assumption is not reasonable. We deal with this technical difficulty as follows. In \( S_{k,0} \) we use a Dirichlet boundary \( \Gamma_0^{(k)} := \cup \{ e_i \mid e_i \in E_k, e_i^{(m)} \cap \Gamma_0 \neq \emptyset \} \) and given boundary values \( \xi_k(x) \) (\( x \in \Gamma_0^{(k)} \)) with \( \xi_k(x) = \xi(x) \) if \( x \in \Gamma_0 \) and \( \xi_k(x) \) equal to a suitable extrapolation if \( x \in \Gamma_0^{(k)} \setminus \Gamma_0 \) (clearly if \( \xi \equiv 0 \) on \( \Gamma_0 \) then \( \xi_k \equiv 0 \) on \( \Gamma_0^{(k)} \)).

For the construction of a multigrid method we need relations between \( S_{k,0} \) for different values of \( k \). We assume that \( T_k \) ("finer triangulation") is obtained from \( T_{k-1} \) ("coarser triangulation") by connecting the midpoints of the edges of the triangles of \( T_{k-1} \).

The prolongation operator (mapping elements in \( S_{k-1,0} \) on elements in \( S_{k,0} \)) will be based on the following Lemma.

Lemma 5.2. Define \( W_k := S_{k,0} \oplus S_{k-1,0} \). Let \( P_k \) be the orthogonal projection w.r.t. the \( L^2 \)-inner product \( \langle \cdot, \cdot \rangle \) of \( W_k \) on \( S_{k,0} \). Take \( u \in W_k \) and take the midpoint \( m_i \) of an edge \( e_i \in E_k \).

If \( e_i \subseteq \partial \Omega \) then \( (P_k u)(m_i) = u(m_i) \). If \( e_i \not\subseteq \partial \Omega \) then \( (P_k u)(m_i) = \left( |T^L| + |T^R| \right)^{-1} \left( |T^L| u_{T^L}(m_i) + |T^R| u_{T^R}(m_i) \right) \), where \( T^L, T^R \subseteq T_k \) are the two triangles with edge \( e_i \).

Proof. If \( e_i \subseteq \Gamma_0^{(k)} \) it is clear that \( (P_k u)(m_i) = u(m_i) \).

Note that \( (P_k u, w) = (u, w) \) for all \( w \in S_{k,0} \). Now take for \( w \) the basis function \( \phi_i \in S_{k,0} \). If \( e_i \subseteq \partial \Omega \setminus \Gamma_0^{(k)} \) then using Lemma 3.1 (e) we get, with \( T_i := \text{supp}(\phi_i) \),

\[
(P_k u, \phi_i) = (u, \phi_i) \Leftrightarrow \int_{T_i} (P_k u) \phi_i \, dx = \int_{T_i} u \phi_i \, dx \Leftrightarrow \frac{T_i}{3} (P_k u)(m_i) = \frac{T_i}{3} u(m_i),
\]

so \((P_k u)(m_i) = u(m_i) \) holds. If \( e_i \not\subseteq \partial \Omega \) then \((P_k u, \phi_i) = (u, \phi_i) \Leftrightarrow \int_{T^L \cup T^R} (P_k u) \phi_i \, dx = \int_{T^L \cup T^R} u \phi_i \, dx \)

\[
\Rightarrow \text{(note that \( P_k u \) is continuous in \( m_i \)) \quad \frac{T^L}{3} (P_k u)(m_i) + \frac{T^R}{3} (P_k u)(m_i) = \frac{T^L}{3} u_{T^L}(m_i) + \frac{T^R}{3} u_{T^R}(m_i), \quad \text{so}
\]

\[
(P_k u)(m_i) = \left( |T^L| + |T^R| \right)^{-1} \left( |T^L| u_{T^L}(m_i) + |T^R| u_{T^R}(m_i) \right).
\]

Corollary 5.3. From Lemma 5.2 it is clear that if \( u \) is continuous at \( m_i \) then \((P_k u)(m_i) = u(m_i)\).

Using Lemma 5.2 it is plain how to generate the representation of \( P_k : S_{k-1,0} \rightarrow S_{k,0} \) with respect to the standard bases in \( S_{k-1,0} \) and \( S_{k,0} \). Let \( M \) be an interior midpoint of \( S_{k-1,0} \) (so \( M \not\subseteq \partial \Omega \)) with corresponding basis function \( \phi \in S_{k-1,0} \), and let \( \bar{T}, \bar{T} \subseteq T_{k-1} \) be such that \( M \in \bar{T} \cap \bar{T} \). We use an enumeration of the basis functions \( \phi_i \in S_{k,0} \) as indicated in Fig. 2 below.
Fig. 2.

Note that $\phi(m_1) = \phi(M) = \phi(m_4) = 1$, so $\phi$ is discontinuous only in $m_i$ with $i \geq 9$.
Define $a_i := |\text{supp}(\phi_i)|$, $\alpha_i := |\text{supp}(\phi_i) \cap (\hat{T} \cup \hat{T})|$.
Using Lemma 5.2 and Corollary 5.3 we get:

$$p_k \phi = \phi_1 + \phi_4 + \frac{1}{2} \sum_{i=2}^{6} \phi_i + \frac{1}{2} \sum_{i=9}^{12} \alpha_i \phi_i - \frac{1}{2} \sum_{i=13}^{16} \alpha_i \phi_i .$$

Now using the above prolongation operator, the multigrid method considered in [3], [5] runs as follows.
Algorithm. (On level $k$ for finding $u_k^{*} \in S_{k,0}$ such that $a_k(u_k^{*}, \phi) = \delta_k(\phi) \; \forall \phi \in S_{k,0}$)

1. **Pre-smoothing.** Given $u_k^{0}$ apply $\nu_1$ smoothing iterations resulting in $u_k^{(\nu_1)}$.

2. **Coarse grid correction.** Let $u_{k-1}^{*} \in S_{k-1,0}$ be the solution of:
   
   $a_{k-1}(u_{k-1}, \phi) = \delta_k(P_k \phi) - a_k(u_k^{(\nu_1)}, P_k \phi)$ for all $\phi \in S_{k-1,0}$.
   
   (cgc)
   
   If $k = 1$ then compute $\tilde{u}_{k-1} := u_{k-1}^{*}$. If $k > 1$ then compute an approximation $\tilde{u}_{k-1}$ of $u_{k-1}^{*}$ by applying $\mu = 1$ or $\mu = 2$ iterations of the algorithm at level $k-1$ for solving (cgc) with starting vector 0.
   
   Put $u_k^{\text{new}} = u_k^{(\nu_1)} + P_k \tilde{u}_{k-1}$.

3. **Post-smoothing.** Apply $\nu_2$ smoothing iterations.

**Remark 5.4.** For this algorithm, applied to (9), convergence proofs can be found in [3], [5] (for the case $\Gamma_0 = \partial \Omega, \gamma = 0$). In particular results are obtained there for the $W$-cycle ($\mu = 2$) if the number of smoothing iterations (damped Jacobi) is large enough. In [3] one can also find a result for the two-grid iteration with only one smoothing step. In order to get this result the algorithm above was modified using a suitable steplength parameter.

**Remark 5.5.** Using the usual isomorphism $I_I : H_{\text{dim}}(S_{I,0}) \rightarrow S_{I,0}$ between functions
in $S_{l,0}$ and the corresponding coefficient vectors, one easily verifies that the fine- and coarse grid equations in (9) and (cgc) can be represented as follows (use the notation $\tilde{u}_t := I^{-1}_t u_t$ for $u_t \in S_{l,0}$).

(9'): $L_k \tilde{u}_t^* = b_k$ with $(b_k)_i = \tilde{g}_k(\phi_i) = -(L_k \tilde{u}_t^d)_i + (g, \phi_i)$

(cgc'): $L_{k-1} \tilde{u}_t^* = r_k (b_k - L_k \tilde{u}_k^d)$

here $L_t$ is the stiffness matrix corresponding to $a_t(\cdot, \cdot)$ and $r_k = p_k^* := (I_k^{-1} P_k I_k)\ast$ (adjoint w.r.t. Euclidean inner product).

Note that in this nonconforming situation $L_{k-1} \neq r_k L_k p_k$.

6. Multigrid method for the rescaled Lagrange multiplier system

In practice we are not able to solve the system (6B) due to the large range of the exponentials. To overcome this problem we rescale the Lagrange multiplier $\lambda_k$, which is an approximation of the Slotboom variable, back to the original variable $u = \rho e^{-\psi}$.

On account of Theorem 4.1 this can be done both for the problem (6A,B) and for the problem (7A,B). We use the latter problem because, due to the nonconforming finite elements used there, it is better suited for developing a multigrid solver based on the multigrid solver of §5.

Let $S_k$ (with basis $\{\phi_i\}_{i \in I \cup I_0}$) and $S_{k,\xi}$ be as in §4. We define the linear operator $\tilde{Q}_k : S_k \rightarrow S_k$ by $\tilde{Q}_k \phi_i = e^{\xi_i} \phi_i$ ($i \in I \cup I_0$).

Before rewriting (7A) we commit a "variational crime" by replacing $g$ (which occurs in $\chi = e^{\psi} g$ in (7A) (and in (6A)) by $\tilde{g}$ with $\tilde{g}(m_i) = g(m_i)$ and $\tilde{g}|_{e_i} \in P_0(e_i)$ for $e_i \in \Gamma_0$.

With this modification $\tilde{Q}_k$ is an isomorphism $S_{k,\tilde{g}} \rightarrow S_k,\chi$.

Now rewriting (7A) with $\tilde{u}_k := Q_k^{-1} \tilde{\eta}_k$ results in

\[
\text{Find } \tilde{u}_k \in S_{k,\tilde{g}} \text{ such that } \left\{ \sum_T \int_T (e^{\psi}|_T)^{-1} \nabla(\tilde{Q}_k \tilde{u}_k) \cdot \nabla \phi \, dx = -\int\nabla \tilde{G}_k(f) \phi \, dx \text{ for all } \phi \in S_{k,0} \right. \]

One additional space is introduced, namely $S_k$ without continuity conditions: $\tilde{S}_k := \{v \in L^2(\Omega) \mid \nabla v \text{ is linear for all } T \in T_k\}$.

Define $Q_k : S_k \rightarrow \tilde{S}_k$ by $(Q_k u)|_T = (e^{\psi}|_T)^{-1}(Q_k u)|_T$ (note that $Q_k$ has a natural extension to $\tilde{S}_k$).

Using this $Q_k$ and the (nonsymmetric) bilinear from $a^Q_k(u,v) := a_k(Q_k u, v) = \sum_T \int_T \nabla (Q_k u) \cdot \nabla v \, dx$ (cf §5) we can rewrite (10) as follows:

\[
\text{Find } u_k^* \in S_{k,0} \text{ such that } \left\{ a^Q_k(u_k^*, \phi) = -a^Q_k(u_k^d, \phi) - \int\nabla \tilde{G}_k(f) \phi \, dx \text{ for all } \phi \in S_{k,0} \right. \]

with $u_k^d := \sum_{i \in E_0} \tilde{g}|_{e_i} \phi_i$.  

11
Remark 6.1. The stiffness matrix corresponding to (F) can be obtained from the stiffness matrix of (6B) (or 7B) by multiplying each column by the corresponding scaling factor $e^Ple_\iota$. So the stiffness matrix of (F) is of the form $AD$ with $A$ an $M$-matrix and $D$ a positive definite diagonal matrix. It is easy to prove that $AD$ is an $M$-matrix, using the following criterion (cf. [11]): Let $C$ be a matrix with $C_{ii} > 0$ for all $i$ and $C_{ij} \leq 0$ for all $i, j$ with $i \neq j$. Let $U := \text{diag}(C) - C$. Then $C$ is an $M$-matrix iff $\rho((\text{diag}(C))^{-1}U) < 1$.

The problem (F) is the final problem we want to solve. Note that if $k \to \infty$ then $Q_k \to \text{Identity}$ so $a_k^Q \approx a_k$ for $k$ large enough (this is not surprising because the highest order term in (2) corresponds to the bilinear form $a_k$). Furthermore if $\psi$ is constant on $T$ then $(Q_k u)_{|T} = u$. Often in semiconductor problems $\psi$ is almost constant in large parts of the domain.

These observations suggest that for developing a solver for (F) it is sensible to use as a starting-point a solver which is suited for solving (F) in the case $Q_k = \text{Identity}$. The multigrid method of §5 meets this requirement.

If we want to apply the algorithm of §5 to the problem (F) we have to specify bilinear forms $a_l^Q(u, v)$ for $0 \leq l < k$ ($u, v \in S_{l,0}$), and we must make a choice for the smoother. We discuss these two things below:

**Coarse grid approximation.** Looking at the bilinear form in the given problem (F) on level $k$, the following two possibilities are obvious:

1. $a_l^Q(u,v) = a_k(Q_k u,v)$ \ (u,v \in S_{l,0})

2. $a_l^Q(u,v) = a_k(Q_k u,v) = a_l(Q_k u,v)$ \ (u,v \in S_{l,0}).

Both possibilities have their own draw-back(s). In 1 the bilinear form $a_k(\cdot, \cdot)$ is acting on $Q_k S_{l,0} \times S_{l,0}$, which leads to a kind of Petrov-Galerkin approximation; we were not even able to prove that this bilinear form leads to uniquely solvable problems. In 2 we get a bilinear form which is an analogon of the one on level $k$, so we get an $M$-matrix; however there are large differences between $Q_l$ and $Q_{l+1}$ for $l < k$ (cf. the definition of $Q_l$: the average exponential over the edges of a triangle in $T_l$ is divided by the average exponential over the whole triangle) which may lead to a violation of the "approximation property" (cf. [7]) or even to difficulties due to overflow.

**Smoothing** An obvious possibility is Gauss-Seidel. However it is known (cf. [7]) that ILU is an excellent smoother for convection diffusion problems with strong convection, so this smoother is a serious candidate too.

For completeness we copy the algorithm of §5 for the situation here:

**Algorithm** (on level $k$, for finding $u_k^* \in S_{k,0}$ such that $a_k^Q(u_k^*, \phi) = g_k(\phi)$ \ \forall \phi \in S_{k,0})

1. **Pre-smoothing.** Given $u_k^Q$ apply $\nu_1$ smoothing iterations resulting in $u_k^{(\nu_1)}$.

2. **Coarse grid correction.** Let $u_{k-1}^* \in S_{k-1,0}$ be the solution of $a_{k-1}^Q(u_{k-1}, \phi) = g_k(P_k \phi) - a_k^Q(u_k^{(\nu_1)}, P_k \phi)$ for all $\phi \in S_{k-1,0}$.
If $k = 1$ then compute $\tilde{u}_{k-1} := u_{k-1}^*$. If $k > 1$ then compute an approximation $\tilde{u}_{k-1}$ of $u_{k-1}^*$ by applying $\mu = 1$ or $\mu = 2$ iterations of the algorithm at level $k - 1$ for solving $(\text{cgc})$ with starting vector $0$.

Put $u_k^{\text{new}} = u_k^{(\nu_2)} + P_k \tilde{u}_{k-1}$.

3. *Post-smoothing.* Apply $\nu_2$ smoothing iterations.

**Remark 6.2.** It is clear that further investigations concerning the choice of smoothing and coarse grid approximation are needed. Numerical experiments (that will be published in a forthcoming paper) with coarse grid approximation 2 and with a suitable Gauss-Seidel smoother yield grid independent convergence for all the test problems in [6].

Two other approaches, concerning the coarse grid approximation, we want to mention are based on the given matrix $L_k$ on the highest level $k$ (so we drop the variational setting).

One might compute $L_l$ for $l < k$ by $L_l := rL_{l+1}P$ with $r = p^*$ and $p = p_{l+1}$ is based on the prolongation $P_{l+1}$ of $\S$ (cf. Remark 5.5). However this is not a good choice because it results in very large difference stars for $L_l$.

Another possibility is to compute $L_l$ for $l < k$ by $L_l := rL_{l+1}P$ but now use matrix dependent prolongations and restrictions as recommended for convection-diffusion problems in [7].

**References**


## Previous Publications in This Series:

<table>
<thead>
<tr>
<th>Number</th>
<th>Author(s)</th>
<th>Title</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>90-01</td>
<td>A. Reusken</td>
<td>Steplength optimization and linear multigrid methods</td>
<td>January '90</td>
</tr>
<tr>
<td>90-02</td>
<td>J. de Graaf</td>
<td>On approximation of (slowly growing) analytic functions by special polynomials</td>
<td>February '90</td>
</tr>
<tr>
<td>90-03</td>
<td>A.A.F. van de Ven and P.H. van Lieshout</td>
<td>Buckling of superconducting structures under prescribed current</td>
<td>February '90</td>
</tr>
<tr>
<td>90-04</td>
<td>F.J.L. Martens</td>
<td>A representation of $GL(q, \mathbb{R})$ in $L_2(S^{q-1})$</td>
<td>April '90</td>
</tr>
<tr>
<td>90-05</td>
<td>J. de Graaf</td>
<td>Skew-Hermitean representations of Lie algebras of vectorfields on the unit-sphere</td>
<td>April '90</td>
</tr>
<tr>
<td>90-06</td>
<td>Y. Shindo and K. Horiguchi and A.A.F. van de Ven</td>
<td>Bending of a magnetically saturated plate with a crack in a uniform magnetic field</td>
<td>April '90</td>
</tr>
<tr>
<td>90-07</td>
<td>M. Kuipers and A.A.F. van de Ven</td>
<td>Unilateral contact of a springboard and a fulcrum</td>
<td>July '90</td>
</tr>
<tr>
<td>90-08</td>
<td>P.H. van Lieshout and A.A.F. van de Ven</td>
<td>A variational approach to the magneto-elastic buckling problem of an arbitrary number of superconducting beams</td>
<td>July '90</td>
</tr>
<tr>
<td>90-09</td>
<td>A. Reusken</td>
<td>A multigrid method for mixed finite element discretizations of current continuity equations</td>
<td>August '90</td>
</tr>
</tbody>
</table>