Weierstrass Institute for
Applied Analysis and Stochastics

Berlin Mathematical School

Study of Iterative Methods for Nonlinear AFC Discretizations of Convection-Diffusion Equations

Abhinav Jha (BMS Berlin and WIAS Berlin), Volker John (WIAS Berlin and Freie Universität Berlin)

Outline of the talk

1 Algebraic Flux Correction Schemes

2 Iteration Schemes

3 Numerical Studies at the 2d Hemker Example

4 Numerical Studies for a Smooth Example

5 Summary and Outlook

- steady-state convection-diffusion-reaction equation

$$\begin{aligned} -\varepsilon\Delta u + \boldsymbol{b}\cdot\nabla u + cu &= g && \text{in } \Omega \\ u &= u^b && \text{on } \partial\Omega_{\mathrm{D}}, \\ -\varepsilon\nabla u\cdot\boldsymbol{n} &= 0 && \text{on } \partial\Omega_{\mathrm{N}} \end{aligned}$$

  - Ω – bounded polyhedral Lipschitz domain in $\mathbb{R}^d$, $d \in \{2,3\}$
  - $\boldsymbol{n}$ – outward pointing unit normal
  - interested in convection-dominated regime $\varepsilon \ll \|\boldsymbol{b}\|$

- ideal discretization
  1. accurate and sharp layers
     - many discretizations satisfy this property, e.g., SUPG
     - reasonably well for AFC schemes
  2. physically consistent results (no spurious oscillations)
     - most discretizations violate this property, e.g., SUPG, SOLD schemes
     - satisfied for AFC schemes
  3. efficient computation of the solutions
     - satisfied for linear discretizations
     - usually not satisfied for nonlinear discretizations, like AFC schemes

- because of 2^{nd} property: AFC schemes very well suited for applications
- this talk: present studies with respect to 3^{rd} property

- derivation
  - Galerkin FEM (algebraic form)

$$\begin{aligned}\sum_{j=1}^{N} a_{ij} u_j &= g_i, \quad i = 1, \ldots, M, \\ u_i &= u_i^b, \quad i = M+1, \ldots, N\end{aligned}$$

  - artificial diffusion matrix D

$$d_{ij} = d_{ji} = -\max\{a_{ij}, 0, a_{ji}\} \; \forall\, i \neq j, \quad d_{ii} = -\sum_{i \neq j} d_{ij}$$

  - anti-diffusive fluxes

$$f_{ij} = d_{ij}(u_j - u_i), \quad f_{ij} = -f_{ji}, \quad i, j = 1, \ldots, N$$

- derivation (cont.)
  - solution-dependent coefficients

$$\alpha_{ij} = \alpha_{ji}, \quad i,j = 1,\ldots,N$$

  with

$$\alpha_{ij} \in [0,1]$$

  - final scheme

$$\begin{aligned}\sum_{j=1}^{N} a_{ij}u_j + \sum_{j=1}^{N}(1-\alpha_{ij})d_{ij}(u_j - u_i) &= g_i, \quad i = 1,\ldots,M,\\ u_i &= u_i^b, \quad i = M+1.\ldots,N\end{aligned}$$

- limiters
  - Kuzmin limiter [1]
  - BJK limiter [2]
  - analytical properties in [2,3,4]
  - BJK limiter in general more accurate [4]

[1] Kuzmin: in Proc. Int. Conf. Comput. Meth. for Coupled Problems in Science and Engineering, CIMNE, 2007

[2] Barrenechea, John, Knobloch: M3AS 27, 525–548, 2017

[3] Barrenechea, John, Knobloch: SINUM 54, 2427–2451, 2016

[4] Barrenechea, John, Knobloch, Rankin: SeMA Journal, in press, 2018

- Kuzmin limiter [1], (non-differentiable operations)
  - compute

$$P_i^+ := \sum_{\substack{j=1\\ a_{ji}\le a_{ij}}}^{N} f_{ij}^+,\ P_i^- := \sum_{\substack{j=1\\ a_{ji}\le a_{ij}}}^{N} f_{ij}^-,\ Q_i^+ := -\sum_{j=1}^{N} f_{ij}^-,\ Q_i^- := -\sum_{j=1}^{N} f_{ij}^+,$$

with $f_{ij}^+ = \max\{0, f_{ij}\}$ and $f_{ij}^- = \min\{0, f_{ij}\}$

  - compute

$$R_i^+ := \min\left\{1, \frac{Q_i^+}{P_i^+}\right\}, \quad R_i^- := \min\left\{1, \frac{Q_i^-}{P_i^-}\right\}$$

  - if $a_{ji} \le a_{ij}$, set

$$\alpha_{ij} := \begin{cases} R_i^+ & \text{if } f_{ij} > 0 \\ 1 & \text{if } f_{ij} = 0 \\ R_i^- & \text{if } f_{ij} < 0 \end{cases} \qquad \alpha_{ji} := \alpha_{ij}$$

[1] Kuzmin: in Proc. Int. Conf. Comput. Meth. for Coupled Problems in Science and Engineering, CIMNE, 2007

- BJK limiter [1]
  - set for appropriate index set S_i and sufficiently large value γ_i

$$u_i^{\max} := \max_{j\in S_i\cup\{i\}} u_j, \quad u_i^{\min} := \min_{j\in S_i\cup\{i\}} u_j, \quad q_i = \gamma_i \sum_{j\in S_i} d_{ij}$$

  - compute

$$P_i^+ := \sum_{j\in S_i} f_{ij}^+,\ P_i^- := \sum_{j\in S_i} f_{ij}^-,\ Q_i^+ := q_i(u_i - u_i^{\max}),\ Q_i^- := q_i(u_i - u_i^{\min})$$

  - compute

$$R_i^+ := \min\left\{1, \frac{Q_i^+}{P_i^+}\right\}, \quad R_i^- := \min\left\{1, \frac{Q_i^-}{P_i^-}\right\}$$

  - set

$$\bar{\alpha}_{ij} := \begin{cases} R_i^+ & \text{if } f_{ij} > 0 \\ 1 & \text{if } f_{ij} = 0 \\ R_i^- & \text{if } f_{ij} < 0 \end{cases}, \quad \alpha_{ij} := \min\{\bar{\alpha}_{ij}, \bar{\alpha}_{ji}\}$$

[1] Barrenechea, John, Knobloch: M3AS 27, 525–548, 2017

- given iterate $u^{(m)}$
- fixed point iteration with changing matrix

$$\sum_{j=1}^{N} a_{ij}\, \tilde{u}_j^{(m+1)} + \sum_{j=1}^{N} \left(1 - \alpha_{ij}^{(m)}\right) d_{ij} \left(\tilde{u}_j^{(m+1)} - \tilde{u}_i^{(m+1)}\right) = g_i,$$
$$\tilde{u}_i^{(m+1)} = u_i^b$$

- fixed point iteration with fixed matrix: using

$$\sum_{j=1}^{N}(1-\alpha_{ij})d_{ij}(u_j - u_i) = \sum_{j=1}^{N} d_{ij}u_j - u_i \underbrace{\sum_{j=1}^{N} d_{ij}}_{=0} - \sum_{j=1}^{N} \alpha_{ij}d_{ij}(u_j - u_i),$$

gives

$$\begin{aligned} \sum_{j=1}^{N}(a_{ij}+d_{ij})\tilde{u}_j^{(m+1)} &= g_i + \sum_{j=1}^{N} \alpha_{ij}^{(m)} f_{ij}^{(m)}, \quad && i = 1,\ldots,M, \\ \tilde{u}_i^{(m+1)} &= u_i^b, && i = M+1,\ldots,N \end{aligned}$$

- fixed point iterations
  - fixed point iteration with fixed matrix
    - matrix is M-matrix
    - with direct sparse solver: factorization only once needed
  - fixed point iteration with changing matrix
    - more implicit approach, hope for better convergence properties
  - general fixed point iteration by linear combination

$$\begin{aligned} & \sum_{j=1}^{N} (a_{ij} + d_{ij})\, \tilde{u}_j^{(m+1)} - \omega_{\mathrm{fp}} \sum_{j=1}^{N} \alpha_{ij}^{(m)} d_{ij} \left(\tilde{u}_j^{(m+1)} - \tilde{u}_i^{(m+1)} \right) \\ & \quad = g_i + (1 - \omega_{\mathrm{fp}}) \sum_{j=1}^{N} \alpha_{ij}^{(m)} f_{ij}^{(m)}, \quad i = 1, \ldots, M, \\ \tilde{u}_i^{(m+1)} & = u_i^b, \qquad\qquad i = M+1, \ldots, N \end{aligned}$$

- formal Newton method
  - formal derivation of Jacobian

$$DF\left(\underline{u}^{(m)}\right)_{ij} = \begin{cases} a_{ij} + d_{ij} - \alpha_{ij}^{(m)} d_{ij} - \sum_{k=1}^{N} \frac{\partial \alpha_{ik}^{(m)}}{\partial u_j} d_{ik} \left(u_k^{(m)} - u_i^{(m)}\right) & \text{if } i \neq j, \\ a_{ii} + d_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{N} \alpha_{ij}^{(m)} d_{ij} - \sum_{k=1}^{N} \frac{\partial \alpha_{ik}^{(m)}}{\partial u_i} d_{ik} \left(u_k^{(m)} - u_i^{(m)}\right) & \text{if } i = j \end{cases}$$

- formal Newton method: how to deal with non-smooth cases?
- discussion only for Kuzmin limiter
  - involves maxima and minima of two arguments, one of them is constant

1. non-regularized approach
   - take one-sided derivative w.r.t. constant, i.e., set value to zero
2. regularized approach
   - replace maximum for some $\sigma > 0$ by [1]

$$\max_\sigma(x,y) = \frac{1}{2}\left(x + y + \sqrt{(x-y)^2+\sigma}\right)$$

   - we did not regularized the limiter in the equation, only in the iteration matrix, since
     - in our opinion: solution should not depend on solver
     - analytical results from literature not longer applicable

[1] Badia, Bonilla: CMAME 313, 133–158, 2017

- general form of the matrix

$$\underbrace{\underbrace{\underbrace{a_{ij} + d_{ij}}_{\text{fp, const. matrix}} \quad -\omega_{\text{fp}}\alpha_{ij}d_{ij}}_{\text{fp, changing matrix}} + \omega_{\text{jac}}(\text{term with der. of } \alpha_{ij}),}_{\text{formal Newton}} \quad i \neq j$$

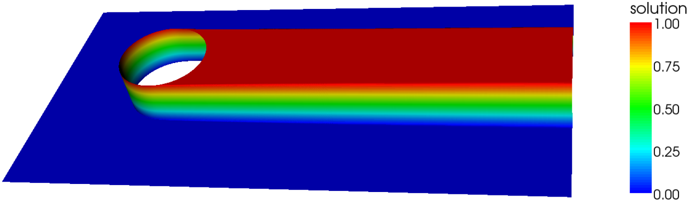
  - similar for diagonal entries
  - some modifications for regularized Newton approach
- iteration

$$\underline{u}^{(m+1)} = \underline{u}^{(m)} + \omega^{(m)} \left(\underline{\tilde{u}}^{(m+1)} - \underline{u}^{(m)} \right)$$

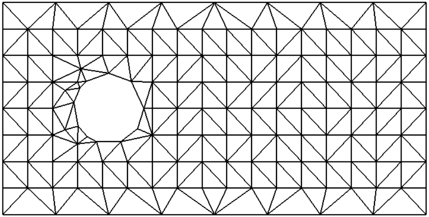
  - adaptive choice of damping parameter as proposed in [1]

[1] John, Knobloch: CMAME 197, 1997–2014, 2008

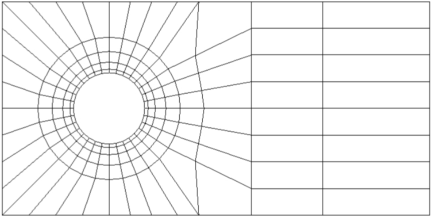
- various values of ε, $\boldsymbol{b} = (1,0)^T$, $c = g = 0$



- standard benchmark problem
- P_1 and Q_1 finite elements
- stop of the iteration
  - $\|\text{residual}\|_2 \le \sqrt{\#\ \text{dof}}\ 10^{-10}$
  - 25000 iterations

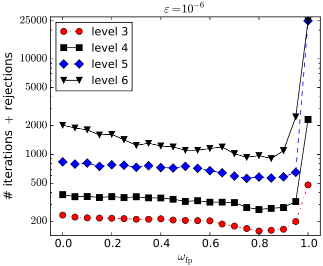


P_1 grid for Hemker example



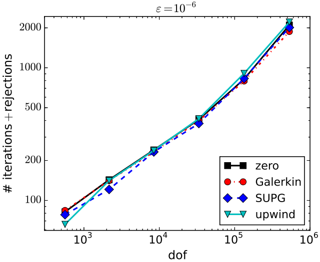
Q_1 grid for Hemker example

- Kuzmin limiter, P_1
- general fixed point iteration, $\varepsilon = 10^{-6}$

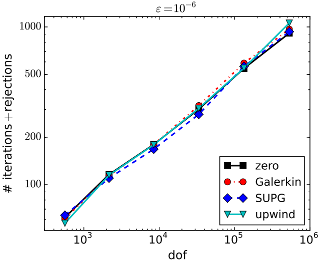


  - number of iterations increases with level
  - $\omega_{\mathrm{fp}} = 0$: method that changes only right-hand side
  - very slow or even no convergence for method which changes only matrix $(\omega_{\mathrm{fp}} = 1)$
  - good parameter in general fixed point iteration is $\omega_{\mathrm{fp}} = 0.75$

- Kuzmin limiter, P_1, dependency on initial iterate
- general fixed point iteration, $\varepsilon = 10^{-6}$
  - zero in all degrees of freedom
  - Galerkin FEM
  - upwind FEM
  - SUPG



$\omega_{\mathrm{fp}} = 0$

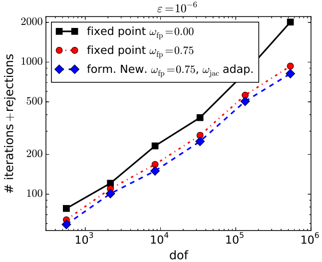


$\omega_{\mathrm{fp}} = 0.75$

  - only minor differences

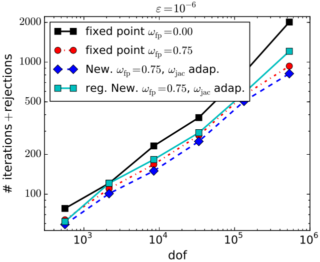
- similar observations for other small values of ε
- summary so far: slow convergence for fixed point iteration which changes only matrix
  - expectation that damping of formal Newton term also necessary: ω_{jac}
  - preliminary tests showed that appropriate value depends on refinement level
  - that's why: simple adaptive choice based on the value of the reduction of the norm of the residual
  - formal Newton term only activated if norm of residual is small ($\leq 10^{-5}$)

- Kuzmin limiter, P_1, start with SUPG solution
- formal Newton method: $\omega_{\mathrm{fp}} = 0.75$, ω_{jac} adaptive



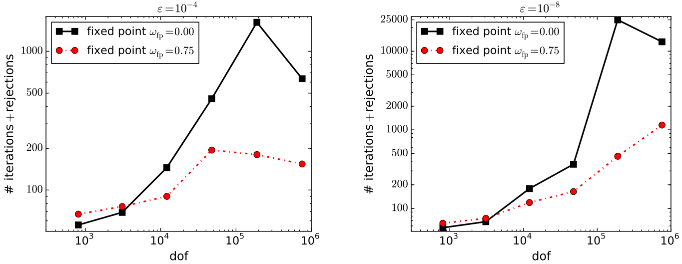
  - only minor reduction of number of iterations compared with general fixed point iteration with $\omega_{\mathrm{fp}} = 0.75$

- Kuzmin limiter, P_1, start with SUPG solution
- formal Newton method with regularization of minima: $\omega_{\mathrm{fp}} = 0.75$, ω_{jac} adaptive



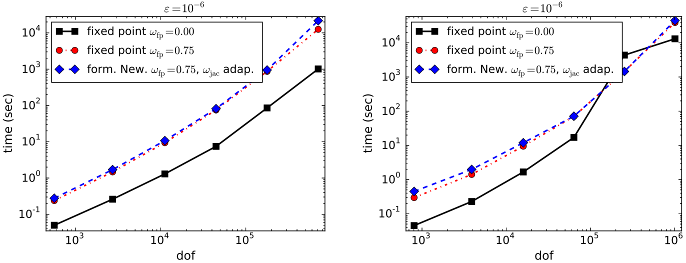
  - no improvement, even more iterations than other methods

- Kuzmin limiter, Q_1, start with SUPG solution
- general fixed point iteration



  - on finer levels: considerably less iterations with $\omega_{\mathrm{fp}} = 0.75$ than with $\omega_{\mathrm{fp}} = 0$
  - no convergence with $\omega_{\mathrm{fp}} = 0$ for $\varepsilon = 10^{-8}$ on level 5 (even not after 100000 iterations)

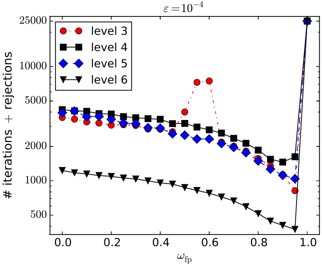
- efficiency in terms of computing time
- Kuzmin limiter, start with SUPG solution
- direct sparse solver (UMFPACK)

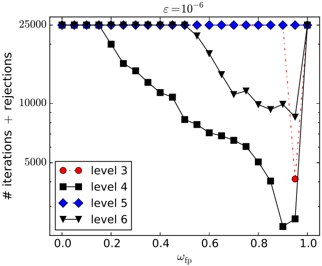


P_1 Q_1

- fixed point iteration with $\omega_{\mathrm{fp}} = 0$ generally the best (often one order of magnitude)
  - factorization of matrix only once

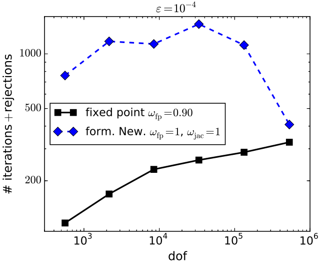
- BJK limiter, P_1
- general fixed point iteration

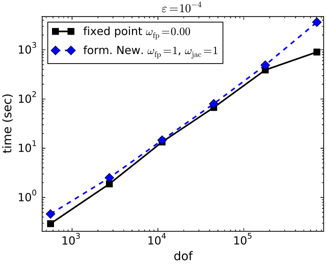




- $\varepsilon = 10^{-4}$
  - good value is $\omega_{\mathrm{fp}} = 0.9$
  - very slow convergence for $\omega_{\mathrm{fp}} = 1$
- $\varepsilon = 10^{-6}$
  - no solver worked on level 5 (also formal Newton did not)

- BJK limiter, P_1, $\varepsilon = 10^{-4}$
- formal Newton method without damping



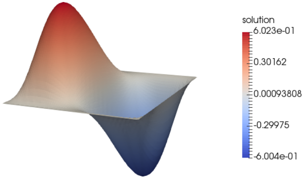


  - much less iterations with formal Newton method (on coarser grids)
  - but no gain in computing time

- formal Newton method with damping much slower than without damping

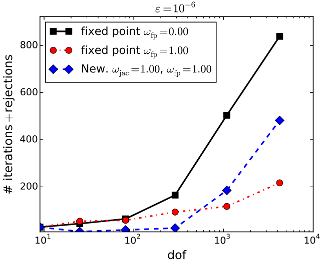
- various values of ε, $\boldsymbol{b} = (3,2)^T$, $c = 1$ and g such that

$$u(x,y) = 10x^2(1-x)^2y(1-y)(1-2y)$$



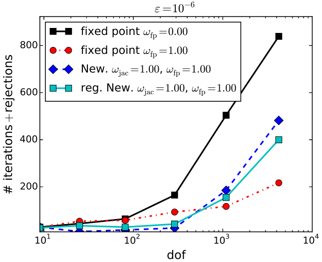
- standard problem having a smooth solution
- P_1 finite elements
- stop of the iteration
  - $\|\text{residual}\|_2 \le \sqrt{\#\text{ dof}}\,10^{-10}$
  - 10000 iterations

- Kuzmin limiter, P_1, start with SUPG solution
- formal Newton method: $\omega_{\text{fp}} = 1.0$, $\omega_{\text{jac}} = 1.0$



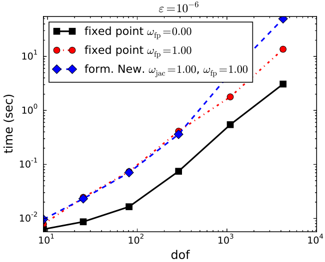
  - only minor reduction in number of iterations compared with fixed point iteration with $\omega_{\text{fp}} = 1.0$ at coarser grids but more iterations at finer grids

- Kuzmin limiter, P_1, start with SUPG solution
- formal Newton method with regularization of minima: $\omega_{\text{fp}} = 1.0$, $\omega_{\text{jac}} = 1.0$



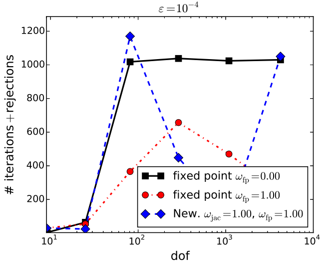
  - little improvement at finer grids, but more iterations than $\omega_{\text{fp}} = 1.0$

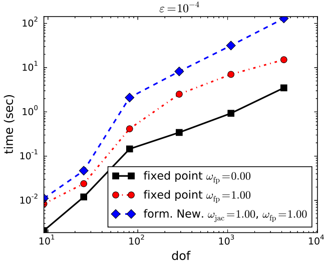
- efficiency in terms of computing time
- Kuzmin limiter, start with SUPG solution
- direct sparse solver (UMFPACK)



  - at coarser grids all the methods perform equally
  - fixed point iteration with $\omega_{\text{fp}} = 0.0$ is better at finer grids
    - factorization of matrix only once

- BJK limiter, P_1, $\varepsilon = 10^{-4}$
- formal Newton method without damping





  - no unified method for all grids
  - $\omega_{\rm fp} = 1.0$ on average gives better results than $\omega_{\rm fp} = 0.0$ and Newton method
- fixed point with $\omega_{\rm fp} = 0.0$ has the best computing times

5 Summary and Outlook

- studied several methods for solving nonlinear problems arising in AFC schemes
  - fixed point iteration with change of right-hand side
  - general fixed point iteration with change of matrix and right-hand side
  - formal Newton-type method (based on formal derivation)
  - regularization of formal Newton-type method (only for Kuzmin limiter)
- observations
  - with appropriate parameters there is some gain w.r.t. the number of iterations for more complicated methods
  - with sparse direct solver: method with constant matrix is usually most efficient
  - easier to solve problems for Kuzmin limiter than for BJK limiter
- altogether: so far disappointing results

5 Summary and Outlook

- outlook
  - other examples to understand behavior of methods better
    - more examples in 2d
    - 3d examples
  - iterative solvers (direct solvers infeasible in 3d)