Weierstrass Institute for
Applied Analysis and Stochastics

Berlin Mathematical School

Investigation of different solvers for Nonlinear Algebraic Stabilizations of Convection-Diffusion Equations

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

Outline of the talk

1 Algebraic Flux Correction (AFC) Schemes

- 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 $\varepsilon \ll \|\boldsymbol{b}\|$

1 Algebraic Flux Correction (AFC) Schemes

- 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 results with respect to the 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 75, 655-685, 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 (FPM)

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

- Fixed point iteration with fixed matrix (FPR): 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 (FPR)
    - Matrix is M-matrix
    - With direct sparse solver: factorization is needed only once
  - Fixed point iteration with changing matrix (FPM)
    - More implicit approach, hope for better convergence properties
  - General fixed point iteration by linear combination

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

- Formal Newton method
  - Formal derivation of Jacobian

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

- 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 are not longer applicable

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

- General form of the matrix

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

  - Similar for diagonal entries
  - Some modifications for regularized Newton approach
- Iteration

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

- Algorithmic components
  - Adaptive choice of damping parameter [1]
  - Anderson acceleration [2]
  - Projection to admissible values [3]
  - Selection of initial iterate

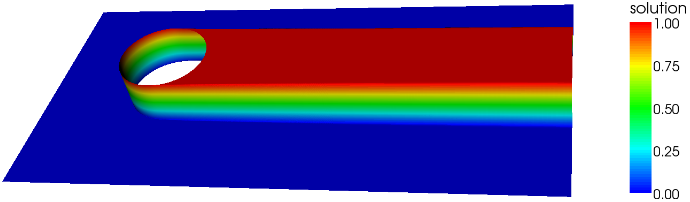
[1] John, Knobloch: CMAME 197, 1997–2014, 2008

[2] Walker, Ni: SINUM 49(4), 1715-1735, 2011

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

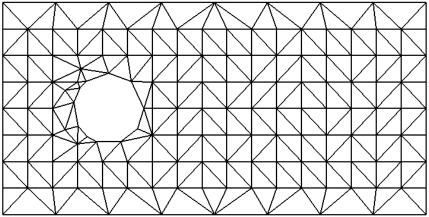
3 Numerical Studies of the 2d Hemker problem

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

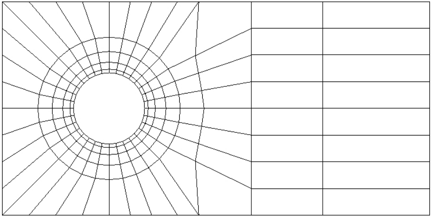


- Standard benchmark problem
- P_1 and Q_1 finite elements
- Stopping criteria
  - $\|\mathrm{residual}\|_2 \le \sqrt{\#\ \mathrm{dof}}\ 10^{-10}$
  - 25000 iterations

3 Numerical Studies of the 2d Hemker problem

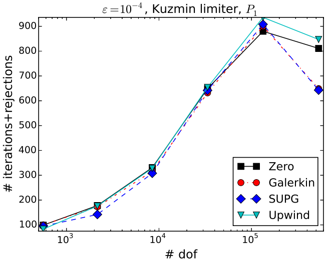


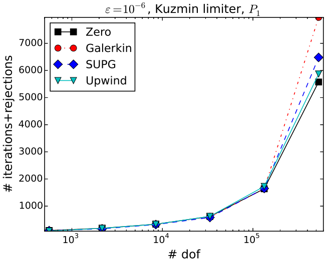
P_1 grid for Hemker example



Q_1 grid for Hemker example

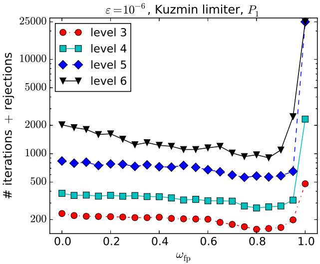
- Kuzmin limiter, P_1, dependency on initial iterate
- General fixed point iteration, $\varepsilon = 10^{-4}, 10^{-6}$
  - Zero in all degrees of freedom
  - Galerkin FEM
  - SUPG
  - Upwind FEM



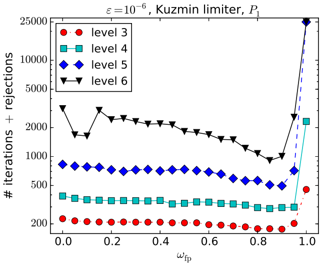


  - Only minor differences

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



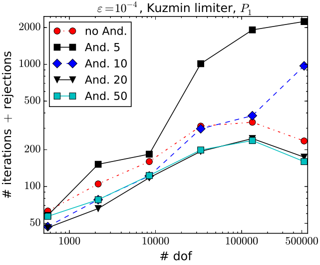
Without projection

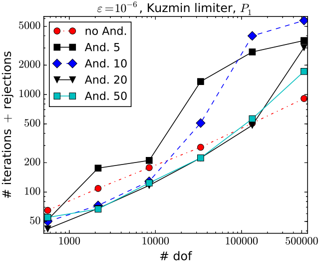


With projection

- Number of iterations increases with level
- $\omega_{\mathrm{fp}} = 0$: FPR method
- Very slow or even no convergence for FPM method ($\omega_{\mathrm{fp}} = 1$)
- Good parameter in general fixed point iteration is $\omega_{\mathrm{fp}} = 0.85$
- Minimal difference after projection to admissible values

- Kuzmin limiter, P_1
- Anderson acceleration with $\omega_{\text{fp}} = 0.85$, $\varepsilon = 10^{-4}, 10^{-6}$





  - $\varepsilon = 10^{-4}$, 20 or 50 Anderson vectors reduced number of iterations
  - $\varepsilon = 10^{-6}$, reduction of iterations only on coarse grids

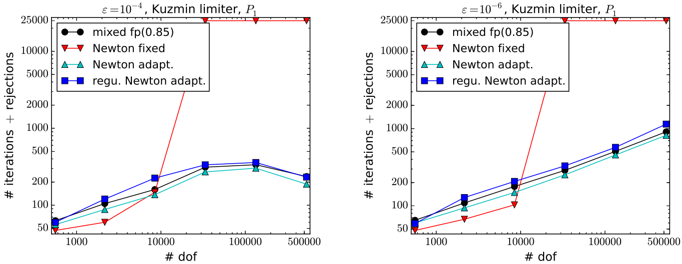
- Similar observations for other small values of ε

- Projection to admissible values don't affect the solution that much

- Summary so far: Slow convergence for FPM method
  - 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 ($\leq 10^{-5}$)

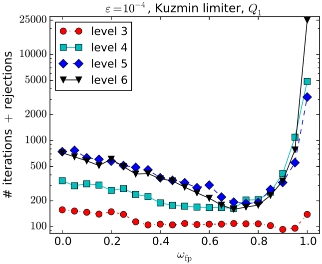
- Kuzmin limiter, P_1, start with SUPG solution
- Formal Newton method: $\omega_{\text{fp}} = 0.85$, ω_{jac} adaptive

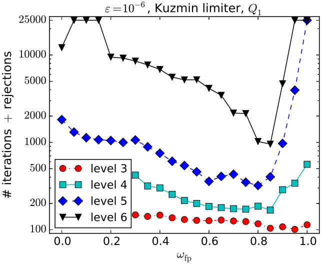


  - Fixed damping reduces iterations only on coarse grid
  - Formal Newton with adaptive ω_{jac} needs less iterations
  - Regularized Newton with adaptive ω_{jac} requires more iterations than $\omega_{\text{fp}} = 0.85$

3 Numerical Studies of the 2d Hemker problem

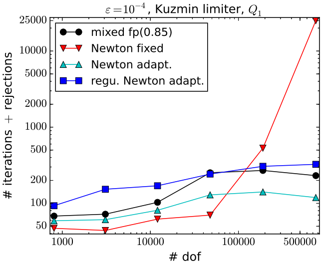
- Kuzmin limiter, Q_1, start with SUPG solution
- General fixed point iteration, $\varepsilon = 10^{-4}, 10^{-6}$

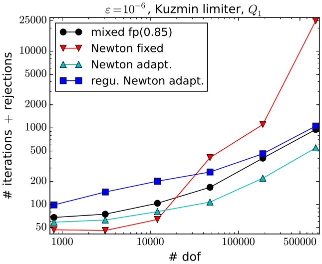




  - Similar observations as P_1 elements

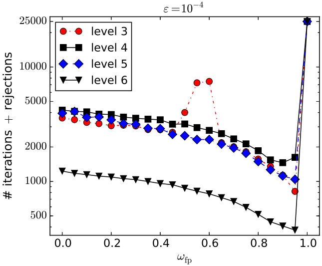
- Kuzmin limiter, Q_1, start with SUPG solution
- Formal Newton method: $\omega_{\text{fp}} = 0.85, \omega_{\text{jac}}$ adaptive

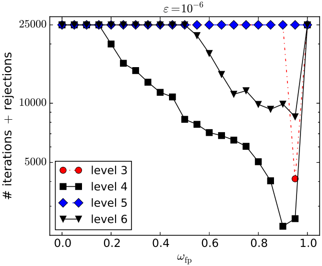




- Similar observations as P_1 elements

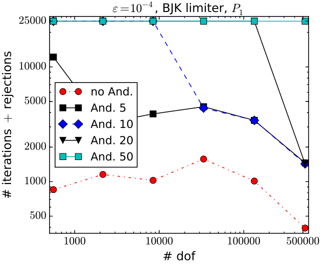
- BJK limiter, P_1
- General fixed point iteration

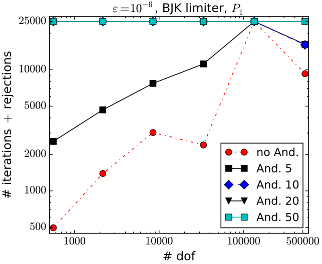




- $\varepsilon = 10^{-4}$
  - Good value is $\omega_{\mathrm{fp}} = 0.95$
  - Very slow convergence for $\omega_{\mathrm{fp}} = 1$
- $\varepsilon = 10^{-6}$
  - No solver worked on level 5

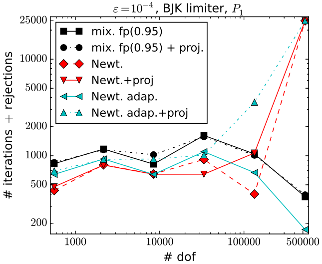
- BJK limiter, P_1
- Anderson acceleration with $\omega_{\mathrm{fp}} = 0.95$, $\varepsilon = 10^{-4}, 10^{-6}$





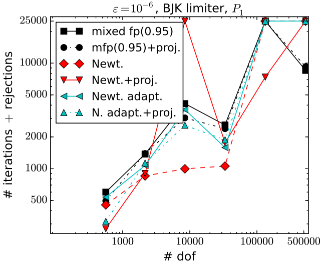
  - Anderson acceleration worsens the convergence for all simulations

- BJK limiter, P_1, $\varepsilon = 10^{-4}$
- Formal Newton method



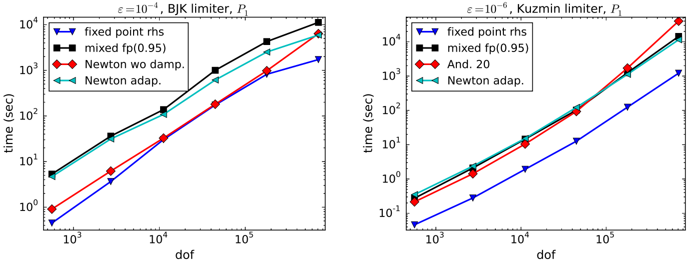
- $\varepsilon = 10^{-4}$
  - Formal Newton with adaptive ω_{jac} needs less iterations without projection
  - Method doesn't converge for fine grids if projection is used

- BJK limiter, P_1, $\varepsilon = 10^{-6}$
- Formal Newton method



  - $\varepsilon = 10^{-6}$
    - Some formal Newton method requires less iteration on coarse grids
    - Some methods behaved differently with and without projection
- No uniform method that works for all cases

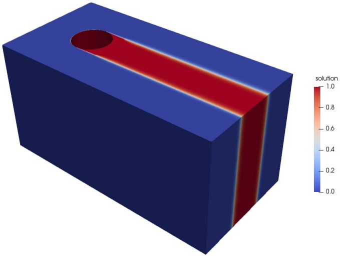
- Efficiency in terms of computing time
- Direct sparse solver (UMFPACK)



- ○ Fixed point iteration with $\omega_{\text{fp}} = 0$ generally the best (often one order of magnitude)
  - Factorization of matrix only once

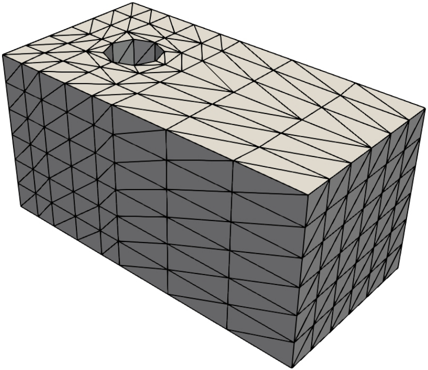
4 Numerical Studies of the 3d Hemker problem

- Extension of the 2d problem with domain
  $\Omega = \{\{(-3,9)\times(-3,3)\}\setminus\{(x,y) : x^2+y^2 \le 1\}\}\times(0,6)$



- P_1 finite elements
- Stopping criteria
  - $\|\mathrm{residual}\|_2 \le \sqrt{\#\,\mathrm{dof}}\,10^{-10}$
  - 25000 iterations
- Non converging iterations after projection to admissible values

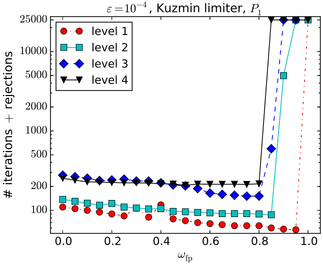
4 Numerical Studies of the 3d Hemker problem

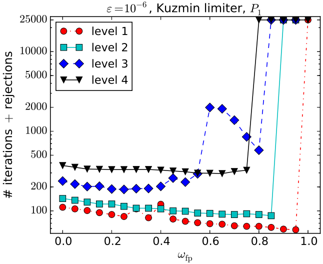


P_1 grid for Hemker example [1]

[1] Wilbrandt et.al, : CAMWA 74, 74-88, 2017

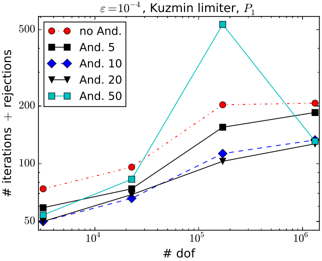
- Kuzmin limiter, P_1
- General fixed point iteration, $\varepsilon = 10^{-4}, 10^{-6}$

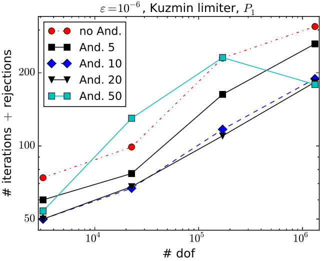




  - Number of iterations increases with level
  - Good parameter in general fixed point iteration is $\omega_{\text{fp}} = 0.5$

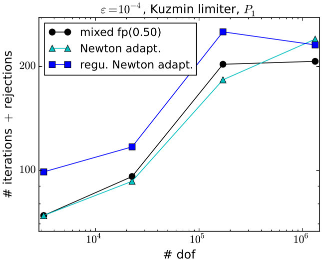
- Kuzmin limiter, P_1
- Anderson acceleration with $\omega_{\text{fp}} = 0.5$, $\varepsilon = 10^{-4}, 10^{-6}$

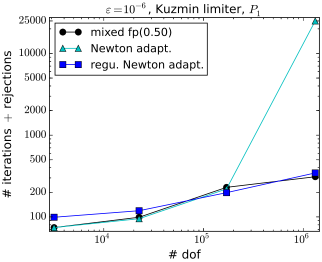




  - 10 or 20 Anderson vectors significantly reduced the number of iterations

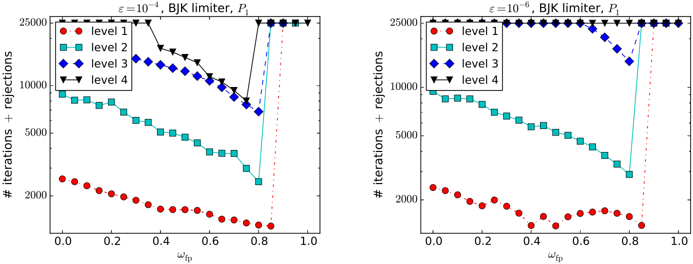
- Kuzmin limiter, P_1, start with SUPG solution
- Formal Newton method: $\omega_{\text{fp}} = 0.5$, ω_{jac} adaptive





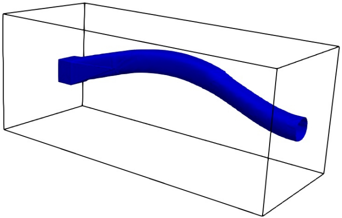
  - Formal Newton method reduced iterations but not considerably.

- BJK limiter, P_1
- General fixed point iteration



  - $\varepsilon = 10^{-4}$
    - Good value is $\omega_{\mathrm{fp}} = 0.7$
  - $\varepsilon = 10^{-6}$
    - No solver worked for finer grids
- Anderson acceleration similar to 2d Hemker problem

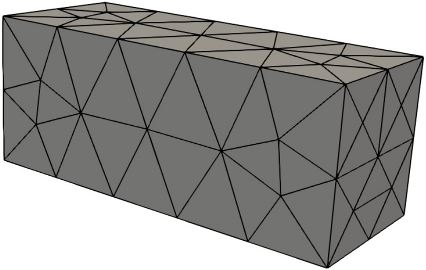
- Various values of ε, $\boldsymbol{b} = (1, l(x), l(x))^T$, $c = g = 0$, where $l(x) = (0.19x^3 - 1.42x^2 + 2.38x)/4$



- Proposed in [1]
- P_1 finite elements
- Stopping criteria
  - $\|\text{residual}\|_2 \leq \sqrt{\#\ \text{dof}}\ 10^{-10}$
  - 25000 iterations

[1] Barrenechea, John, Knobloch, Rankin: SeMA Journal 75, 655–685, 2018

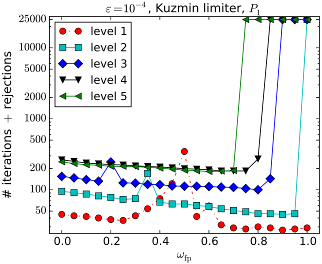
5 Numerical Studies for non-constant convection

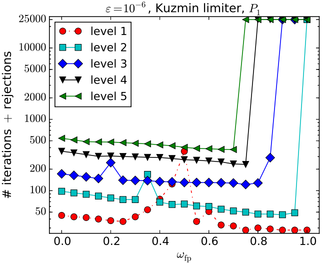


P_1 grid for the example [1]

[1] Geuzaine, Remacle, IJNME 79(11), 1309–131, 2009

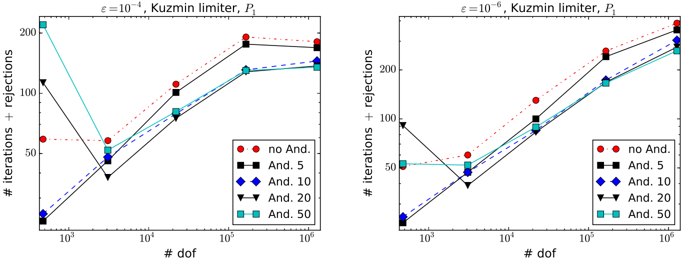
- Kuzmin limiter, P_1
- General fixed point iteration, $\varepsilon = 10^{-4}, 10^{-6}$





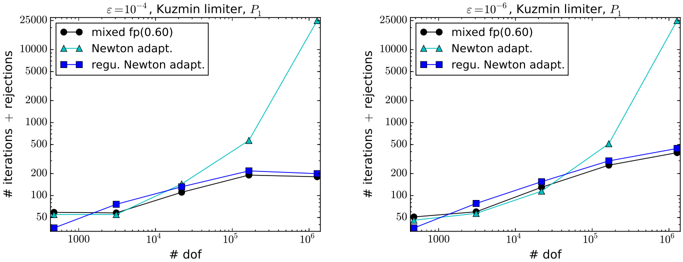
  - Converges for small ω_{fp}
  - Good parameter in general fixed point iteration is $\omega_{\mathrm{fp}} = 0.6$
  - Minimal difference after projection to admissible values

- Kuzmin limiter, P_1
- Anderson acceleration with $\omega_{\mathrm{fp}} = 0.6$, $\varepsilon = 10^{-4}, 10^{-6}$



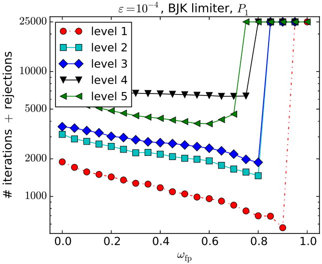
  - 10, 20 or 50 Anderson vectors reduce number of iterations

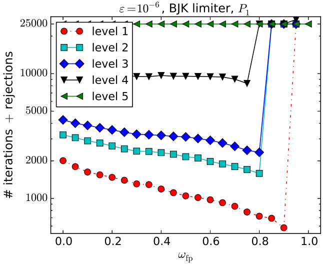
- Kuzmin limiter, P_1, start with SUPG solution
- Formal Newton method: $\omega_{\mathrm{fp}} = 0.6$, ω_{jac} adaptive



  - Formal Newton with adaptive doesn't improve the convergence as compared to $\omega_{\mathrm{fp}} = 0.6$

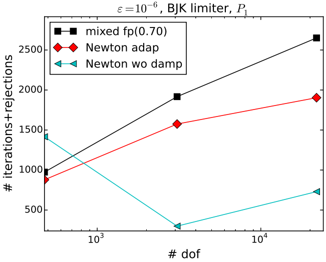
- BJK limiter, P_1
- General fixed point iteration





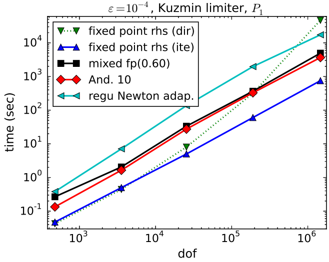
  - Similar behavior to Kuzmin, i.e., small ω_{fp}
  - Needed more iterations as compared to Kuzmin

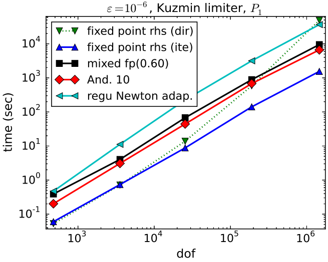
- BJK limiter, P_1
- Formal Newton method



  - Iterative solvers failed, hence direct solvers were used
  - Because of limit on computation only values at three levels were computed
  - Newton without damping reduces significant number of iterations on level 2, 3

- Efficiency in terms of computing time
- Direct sparse solver (UMFPACK) for FPR method
- Iterative solver (GMRES) with preconditioner (SSOR) for other methods





- For fine grids iterative solvers work better
- Fixed point iteration with $\omega_{\text{fp}} = 0$ and iterative solvers takes the least time

6 Conclusion

- 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)
  - Algorithmic components
- Observations
  - With appropriate parameters there is some gain w.r.t. the number of iterations for more complicated methods
  - Easier to solve problems for Kuzmin limiter than for BJK limiter
  - With an appropriate solver:FPR is usually most efficient