# Algebraic stabilizations of convection–diffusion problems and their convergence on general meshes

**Petr Knobloch**

Charles University, Prague

joint work with

**Gabriel R. Barrenechea** (Glasgow)

**Volker John** (Berlin)

10th Workshop Dresden–Prague on Numerical Analysis

Děčín, 4–5 November 2022

# Outline

– discretization of a steady-state convection–diffusion–reaction equation

– algebraic stabilization

– theoretical analysis under general assumptions: solvability, discrete maximum principle, error estimate

– examples of algebraic stabilizations

– investigations of convergence and accuracy on various meshes

– a new algebraically stabilized method

## Steady-state convection–diffusion–reaction equation

$$-\varepsilon\,\Delta u + \mathbf{b}\cdot\nabla u + c\,u = f \quad \text{in } \Omega\,, \qquad u = u_b \quad \text{on } \partial\Omega$$

with constant $\varepsilon > 0$ and

$$\nabla\cdot\mathbf{b} = 0\,, \qquad c \geq \sigma_0 \geq 0 \qquad \text{in } \Omega\,.$$

## Steady-state convection–diffusion–reaction equation

$$-\varepsilon\,\Delta u + \mathbf{b}\cdot\nabla u + c\,u = f \quad \text{in } \Omega\,, \qquad u = u_b \quad \text{on } \partial\Omega$$

with constant $\varepsilon > 0$ and

$$\nabla\cdot\mathbf{b} = 0\,, \qquad c \geq \sigma_0 \geq 0 \qquad \text{in } \Omega\,.$$

**Maximum principle** for any $G \subset \overline{\Omega}$

$$f \leq 0 \quad \text{in } G \qquad \Rightarrow \qquad \max_{\overline{G}} u \leq \max_{\partial G} u^+\,,$$

$$f \geq 0 \quad \text{in } G \qquad \Rightarrow \qquad \min_{\overline{G}} u \geq \min_{\partial G} u^-\,.$$

If $c = 0$ in $G$, then

$$f \leq 0 \quad \text{in } G \qquad \Rightarrow \qquad \max_{\overline{G}} u = \max_{\partial G} u\,,$$

$$f \geq 0 \quad \text{in } G \qquad \Rightarrow \qquad \min_{\overline{G}} u = \min_{\partial G} u\,.$$

## Steady-state convection–diffusion–reaction equation

$$-\varepsilon\,\Delta u + \mathbf{b}\cdot\nabla u + c\,u = f \quad \text{in } \Omega\,, \qquad\qquad u = u_b \quad \text{on } \partial\Omega$$

with constant $\varepsilon > 0$ and

$$\nabla\cdot\mathbf{b} = 0\,, \qquad\quad c \geq \sigma_0 \geq 0 \qquad\qquad \text{in } \Omega\,.$$

## FE discretization

Find $u_h \in W_h$ such that $u_h(x_i) = u_b(x_i)$, $i = M+1,\dots,N$, and

$$a(u_h, v_h) = (f, v_h) \qquad \forall\, v_h \in V_h\,,$$

where

$$W_h = \{v_h \in C(\overline{\Omega})\,;\; v|_K \in P_1(K)\;\forall K \in \mathscr{T}_h\}\,, \qquad V_h = W_h \cap H^1_0(\Omega)\,,$$

$$a(u_h, v_h) = \varepsilon\,(\nabla u_h, \nabla v_h) + (\mathbf{b}\cdot\nabla u_h, v_h) + (c\,u_h, v_h)\,.$$

## Algebraic problem

$$\sum_{j=1}^{N} a_{ij} u_j = f_i \,, \qquad i = 1, \ldots, M \,,$$

$$u_i = u_i^b \,, \qquad i = M+1, \ldots, N \,.$$

## Algebraic problem

$$\sum_{j=1}^{N} a_{ij} u_j = f_i \,, \qquad i = 1, \dots, M \,,$$
$$u_i = u_i^b \,, \qquad i = M+1, \dots, N \,.$$

in the convection-dominated case typically $a_{ij} > 0$ for some $i \neq j$

## Algebraic problem

$$\sum_{j=1}^{N} a_{ij}\,u_j = f_i\,, \qquad i = 1,\dots,M\,,$$
$$u_i = u_i^b\,, \qquad i = M+1,\dots,N\,.$$

in the convection-dominated case typically $a_{ij} > 0$ for some $i \neq j$

$\Rightarrow$ discrete maximum principle (DMP) not satisfied

## Algebraic problem

$$\sum_{j=1}^{N} a_{ij} u_j = f_i \,, \qquad i = 1, \ldots, M \,,$$
$$u_i = u_i^b \,, \qquad i = M+1, \ldots, N \,.$$

in the convection-dominated case typically $a_{ij} > 0$ for some $i \neq j$

$\Rightarrow$ discrete maximum principle (DMP) not satisfied

$\Rightarrow$ approximate solution polluted by spurious oscillations

## Algebraic stabilization

Aim: manipulate the algebraic system in such a way that the solution satisfies DMP and layers are not smeared.

## Algebraic stabilization

Aim: manipulate the algebraic system in such a way that the solution satisfies DMP and layers are not smeared.

Symmetric artificial diffusion matrix $\mathbb{D}$:

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

## Algebraic stabilization

Aim: manipulate the algebraic system in such a way that the solution satisfies DMP and layers are not smeared.

Symmetric artificial diffusion matrix $\mathbb{D}$:

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

## Stabilized algebraic problem

$$\sum_{j=1}^{N} (a_{ij} + d_{ij})\, u_j = f_i\,, \qquad i = 1,\dots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\dots,N\,.$$

## Algebraic stabilization

Aim: manipulate the algebraic system in such a way that the solution satisfies DMP and layers are not smeared.

Symmetric artificial diffusion matrix $\mathbb{D}$:

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

## Stabilized algebraic problem

$$\sum_{j=1}^{N} (a_{ij} + d_{ij})\, u_j = f_i\,, \qquad i = 1, \dots, M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1, \dots, N\,.$$

$\Rightarrow$ DMP holds but excessive smearing of layers occurs

# Nonlinear algebraic problem

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1, \dots, M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1, \dots, N\,.$$

with a general artificial diffusion matrix $\mathbb{B}(\mathrm{U}) = (b_{ij}(\mathrm{U}))_{i,j=1}^{N}$:

$$b_{ij}(\mathrm{U}) = b_{ji}(\mathrm{U})\,, \qquad i, j = 1, \dots, N\,,$$

$$b_{ij}(\mathrm{U}) \leq 0, \qquad i \neq j,\ i, j = 1, \dots, N\,,$$

$$\sum_{j=1}^{N} b_{ij}(\mathrm{U}) = 0\,, \qquad i = 1, \dots, N\,,$$

$$b_{ij}(\mathrm{U}) = 0 \qquad \forall\, j \notin S_i \cup \{i\},\ i = 1, \dots, M\,,$$

where

$S_i = \{j \in \{1, \dots, N\} \setminus \{i\}\,;\ x_i \text{ and } x_j \text{ are endpoints of the same edge}\}$

## Analysis of the nonlinear algebraic problem

Assumption (A1): The functions $b_{ij}$ are bounded on $\mathbb{R}^N$ and the functions $b_{ij}(\mathrm{U})(u_j - u_i)$ are continuous functions of $\mathrm{U} = (u_1, \ldots, u_N)$.

Theorem Under Assumption (A1), the nonlinear algebraic problem is solvable.

## Analysis of the nonlinear algebraic problem

Assumption (A1): The functions $b_{ij}$ are bounded on $\mathbb{R}^N$ and the functions $b_{ij}(\mathrm{U})(u_j - u_i)$ are continuous functions of $\mathrm{U} = (u_1, \ldots, u_N)$.

Theorem Under Assumption (A1), the nonlinear algebraic problem is solvable.

Assumption (A2): Consider any $\mathrm{U} \in \mathbb{R}^N$ and any $i \in \{1, \ldots, M\}$. If $u_i$ is a strict local extremum of $\mathrm{U}$ w.r.t. $S_i$, i.e.,

$$u_i > u_j \quad \forall\, j \in S_i \qquad \mathit{or} \qquad u_i < u_j \quad \forall\, j \in S_i\,,$$

then

$$a_{ij} + b_{ij}(\mathrm{U}) \leq 0 \qquad \forall\, j \in S_i\,.$$

## Analysis of the nonlinear algebraic problem

Theorem Let Assumptions (A1) and (A2) be satisfied. Let $u_h \in W_h$ correspond to the solution of the nonlinear algebraic problem. Consider any nonempty set $\mathscr{G}_h \subset \mathscr{T}_h$ and define $G_h = \cup_{T \in \mathscr{G}_h} T$. Then one has the DMP

$$f \le 0 \quad \text{in } G_h \qquad \Rightarrow \qquad \max_{G_h} u_h \le \max_{\partial G_h} u_h^+ \,,$$

$$f \ge 0 \quad \text{in } G_h \qquad \Rightarrow \qquad \min_{G_h} u_h \ge \min_{\partial G_h} u_h^- \,.$$

If $c = 0$ in $G_h$, then

$$f \le 0 \quad \text{in } G_h \qquad \Rightarrow \qquad \max_{G_h} u_h = \max_{\partial G_h} u_h \,,$$

$$f \ge 0 \quad \text{in } G_h \qquad \Rightarrow \qquad \min_{G_h} u_h = \min_{\partial G_h} u_h \,.$$

# Nonlinear algebraic problem

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1,\ldots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\ldots,N\,.$$

## Nonlinear algebraic problem

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1, \ldots, M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1, \ldots, N\,.$$

## Variational form

Find $u_h \in W_h$ such that $u_h(x_i) = u_b(x_i)$, $i = M+1, \ldots, N$, and

$$a(u_h, v_h) + b_h(u_h; u_h, v_h) = (f, v_h) \qquad \forall\, v_h \in V_h\,,$$

where $$b_h(w; z, v) = \sum_{i,j=1}^{N} b_{ij}(w)\, z(x_j)\, v(x_i)\,.$$

# A priori error estimate

Natural norm: $\|v\|_h = \left(\varepsilon\,|v|_{1,\Omega}^2 + \sigma_0\,\|v\|_{0,\Omega}^2 + b_h(u_h;v,v)\right)^{1/2}$

Theorem Let $u \in H^2(\Omega)$ and $\sigma_0 > 0$. Then

$$\|u-u_h\|_h \le C\left(\varepsilon + \sigma_0^{-1}\{\|\mathbf{b}\|_{0,\infty,\Omega}^2 + \|c\|_{0,\infty,\Omega}^2\}\right)^{1/2} h\,|u|_{2,\Omega} + b_h(u_h; i_h u, i_h u)^{1/2}\,.$$

# A priori error estimate

Natural norm: $$\|v\|_h = \left( \varepsilon\,|v|^2_{1,\Omega} + \sigma_0\,\|v\|^2_{0,\Omega} + b_h(u_h;v,v) \right)^{1/2}$$

Theorem Let $u \in H^2(\Omega)$ and $\sigma_0 > 0$. Then

$$\|u-u_h\|_h \le C\big(\varepsilon + \sigma_0^{-1}\{\|\mathbf{b}\|^2_{0,\infty,\Omega} + \|c\|^2_{0,\infty,\Omega}\}\big)^{1/2} h\,|u|_{2,\Omega} + b_h(u_h;i_h u, i_h u)^{1/2}.$$

Lemma If $|b_{ij}(u_h)| \le \max\{|a_{ij}|,|a_{ji}|\}\ \forall\, i \neq j$, then

$$b_h(u_h;i_h u,i_h u) \le C\,(\varepsilon + \|\mathbf{b}\|_{0,\infty,\Omega}\,h + \|c\|_{0,\infty,\Omega}\,h^2)\,|i_h u|^2_{1,\Omega}.$$

## A priori error estimate

Natural norm: $\|v\|_h = \left(\varepsilon\,|v|_{1,\Omega}^2 + \sigma_0\,\|v\|_{0,\Omega}^2 + b_h(u_h;v,v)\right)^{1/2}$

Theorem Let $u \in H^2(\Omega)$ and $\sigma_0 > 0$. Then

$$\|u-u_h\|_h \le C\big(\varepsilon + \sigma_0^{-1}\{\|\mathbf{b}\|_{0,\infty,\Omega}^2 + \|c\|_{0,\infty,\Omega}^2\}\big)^{1/2} h\,|u|_{2,\Omega} + b_h(u_h;i_h u, i_h u)^{1/2}.$$

Lemma If $|b_{ij}(u_h)| \le \max\{|a_{ij}|,|a_{ji}|\}\ \forall\, i \ne j$, then

$$b_h(u_h;i_h u, i_h u) \le C\,(\varepsilon + \|\mathbf{b}\|_{0,\infty,\Omega}\,h + \|c\|_{0,\infty,\Omega}\,h^2)\,|i_h u|_{1,\Omega}^2\,.$$

Corrollary Let $u \in H^2(\Omega)$ and $\sigma_0 > 0$. Then

$$\|u-u_h\|_h \le C\,h\,\|u\|_{2,\Omega} + C\,(\varepsilon + \|\mathbf{b}\|_{0,\infty,\Omega}\,h)^{1/2}\,|i_h u|_{1,\Omega}\,.$$

## **Algebraic flux correction** Kuzmin et al. (2001–now)

Symmetric artificial diffusion matrix $\mathbb{D}$:

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

Equivalent form of the algebraic problem:

$$\sum_{j=1}^{N} (a_{ij} + d_{ij})\, u_j = f_i + \sum_{j=1}^{N} d_{ij}\, u_j\,, \qquad i = 1, \dots, M$$

## **Algebraic flux correction** Kuzmin et al. (2001–now)

Symmetric artificial diffusion matrix $\mathbb{D}$:

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

Equivalent form of the algebraic problem:

$$\sum_{j=1}^{N} (a_{ij} + d_{ij})\, u_j = f_i + \sum_{j=1}^{N} d_{ij}\, u_j\,, \qquad i = 1, \dots, M$$

Decomposition into fluxes:

$$\sum_{j=1}^{N} d_{ij}\, u_j = \sum_{j\neq i} f_{ij} \quad \text{with} \quad f_{ij} = d_{ij}\,(u_j - u_i).$$

## **Algebraic flux correction** Kuzmin et al. (2001–now)

Symmetric artificial diffusion matrix $\mathbb{D}$:

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

Equivalent form of the algebraic problem:

$$\sum_{j=1}^{N} (a_{ij} + d_{ij})\, u_j = f_i + \sum_{j=1}^{N} d_{ij}\, u_j\,, \qquad i = 1, \dots, M$$

Decomposition into fluxes:

$$\sum_{j=1}^{N} d_{ij}\, u_j = \sum_{j\neq i} f_{ij} \quad \text{with} \quad f_{ij} = d_{ij}\,(u_j - u_i).$$

Limiting: limit those antidiffusive fluxes $f_{ij}$ that would otherwise cause spurious oscillations:

$$\sum_{j=1}^{N} (a_{ij} + d_{ij})\, u_j = f_i + \sum_{j\neq i} \alpha_{ij}\, f_{ij}\,, \qquad i = 1, \dots, M, \quad \alpha_{ij} \in [0, 1]\,.$$

## Algebraic flux correction scheme

$$\sum_{j=1}^{N} a_{ij}\,u_j + \sum_{j=1}^{N} (1-\alpha_{ij})\,d_{ij}\,(u_j - u_i) = f_i\,, \qquad i = 1,\dots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\dots,N\,,$$

where $\alpha_{ij} = \alpha_{ij}(u_1,\dots,u_N) \in [0,1]$,

$$\alpha_{ij} = \alpha_{ji}\,, \qquad i,j = 1,\dots,N\,.$$

# Algebraic flux correction scheme

$$\sum_{j=1}^{N} a_{ij}\,u_j + \sum_{j=1}^{N} \overbrace{(1-\alpha_{ij})\,d_{ij}}^{b_{ij}(\mathrm{U})}(u_j - u_i) = f_i\,, \qquad i = 1,\ldots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\ldots,N\,,$$

where $\alpha_{ij} = \alpha_{ij}(u_1,\ldots,u_N) \in [0,1]$,

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

## Algebraic flux correction scheme

$$\sum_{j=1}^{N} a_{ij}\,u_j + \sum_{j=1}^{N} \overbrace{(1-\alpha_{ij})\,d_{ij}}^{b_{ij}(\mathrm{U})}(u_j - u_i) = f_i\,, \qquad i = 1,\ldots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\ldots,N\,,$$

where $\alpha_{ij} = \alpha_{ij}(u_1,\ldots,u_N) \in [0,1]$,

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

Theoretical analysis:

Barrenechea, John, K., IMAJNA (2015)

Barrenechea, John, K., SINUM (2016)

Barrenechea, John, K., M3AS (2017)

Barrenechea, John, K., Rankin, SeMA J. (2018)

## Kuzmin's limiter

Zalesak (1979), Kuzmin (2007)

$$P_i^+ := \sum_{\substack{j \in S_i \\ a_{ji} \le a_{ij}}} f_{ij}^+ , \qquad Q_i^+ := -\sum_{j \in S_i} f_{ij}^- , \qquad R_i^+ := \min\left\{1, \frac{Q_i^+}{P_i^+}\right\},$$

$$P_i^- := \sum_{\substack{j \in S_i \\ a_{ji} \le a_{ij}}} f_{ij}^- , \qquad Q_i^- := -\sum_{j \in S_i} f_{ij}^+ , \qquad R_i^- := \min\left\{1, \frac{Q_i^-}{P_i^-}\right\}.$$

$$f_{ij} = d_{ij}\,(u_j - u_i)$$

$$\widetilde{\alpha}_{ij} := \begin{cases} R_i^+ & \text{if } f_{ij} > 0\,, \\ 1 & \text{if } f_{ij} = 0\,, \\ R_i^- & \text{if } f_{ij} < 0\,, \end{cases} \qquad \alpha_{ij} = \alpha_{ji} := \begin{cases} \widetilde{\alpha}_{ij} & \text{if } a_{ji} \le a_{ij}\,, \\ \widetilde{\alpha}_{ji} & \text{else.} \end{cases}$$

– DMP holds if $\min\{a_{ij}, a_{ji}\} \le 0 \quad \forall\, i \neq j$

– without the condition $\min\{a_{ij}, a_{ji}\} \le 0$, the DMP does not hold in general

(cf. K., Proceedings of BAIL 2016)

**BJK limiter** Kuzmin (2012), Barrenechea, John, K. (2016)

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

$$P_i^+ := \sum_{j\in S_i} f_{ij}^+\,, \qquad Q_i^+ := q_i\,(u_i - u_i^{\max})\,, \qquad R_i^+ := \min\left\{1, \frac{\mu_i\, Q_i^+}{P_i^+}\right\},$$

$$P_i^- := \sum_{j\in S_i} f_{ij}^-\,, \qquad Q_i^- := q_i\,(u_i - u_i^{\min})\,, \qquad R_i^- := \min\left\{1, \frac{\mu_i\, Q_i^-}{P_i^-}\right\},$$

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

Assumption (A2) always satisfied

$\Rightarrow$ DMP guaranteed for arbitrary meshes!

## Algebraic flux correction scheme

$$\sum_{j=1}^{N} a_{ij}\,u_j + \sum_{j=1}^{N} \overbrace{(1-\alpha_{ij})\,d_{ij}}^{b_{ij}(\mathrm{U})}(u_j - u_i) = f_i\,, \qquad i = 1,\ldots,M\,,$$

with $d_{ij} = -\max\{a_{ij}, 0, a_{ji}\}$.

# Algebraic flux correction scheme

$$\sum_{j=1}^{N} a_{ij}\, u_j + \sum_{j=1}^{N} \overbrace{(1-\alpha_{ij})\, d_{ij}}^{b_{ij}(\mathrm{U})}\, (u_j - u_i) = f_i\,, \qquad i = 1,\dots,M\,,$$

with $d_{ij} = -\max\{a_{ij}, 0, a_{ji}\}$.

$\Rightarrow \quad b_{ij}(\mathrm{U})$ can be written in the form:

$$b_{ij}(\mathrm{U}) = -\max\{(1-\alpha_{ij}(\mathrm{U}))\, a_{ij}, 0, (1-\alpha_{ji}(\mathrm{U}))\, a_{ji}\}$$

## Algebraic flux correction scheme

$$\sum_{j=1}^{N} a_{ij}\, u_j + \sum_{j=1}^{N} \overbrace{(1-\alpha_{ij})\, d_{ij}}^{b_{ij}(\mathrm{U})} (u_j - u_i) = f_i\,, \qquad i = 1,\dots,M\,,$$

with $d_{ij} = -\max\{a_{ij}, 0, a_{ji}\}$.

$\Rightarrow$ $b_{ij}(\mathrm{U})$ can be written in the form:

$$b_{ij}(\mathrm{U}) = -\max\{(1-\alpha_{ij}(\mathrm{U}))\, a_{ij}, 0, (1-\alpha_{ji}(\mathrm{U}))\, a_{ji}\}$$

$\Rightarrow$ the symmetry of $\alpha_{ij}$ not needed any more!

# Nonlinear algebraically stabilized method

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1, \ldots, M\,,$$
$$u_i = u_i^b\,, \qquad i = M+1, \ldots, N\,.$$

where

$$b_{ij}(\mathrm{U}) = -\max\{(1-\alpha_{ij}(\mathrm{U}))\, a_{ij}, 0, (1-\alpha_{ji}(\mathrm{U}))\, a_{ji}\}\,,$$
$$b_{ii}(\mathrm{U}) = -\sum_{j\neq i} b_{ij}(\mathrm{U}) = 0\,,$$
$$\alpha_{ij}(\mathrm{U}) \in [0,1]\,,$$

$\alpha_{ij}(\mathrm{U})(u_i - u_j)$ are continuous functions of $\mathrm{U}$.

## Nonlinear algebraically stabilized method

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1,\dots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\dots,N\,.$$

where

$$b_{ij}(\mathrm{U}) = -\max\{(1-\alpha_{ij}(\mathrm{U}))\, a_{ij}, 0, (1-\alpha_{ji}(\mathrm{U}))\, a_{ji}\}\,,$$

$$b_{ii}(\mathrm{U}) = -\sum_{j\neq i} b_{ij}(\mathrm{U}) = 0\,,$$

$$\alpha_{ij}(\mathrm{U}) \in [0,1]\,,$$

$\alpha_{ij}(\mathrm{U})(u_i - u_j)$ are continuous functions of $\mathrm{U}$.

Theorem Under the above assumptions, the nonlinear method has a solution and satisfies the above error estimate.

# Nonlinear algebraically stabilized method

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1, \ldots, M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1, \ldots, N\,.$$

where

$$b_{ij}(\mathrm{U}) = -\max\{(1-\alpha_{ij}(\mathrm{U}))\, a_{ij}, 0, (1-\alpha_{ji}(\mathrm{U}))\, a_{ji}\}\,,$$

$$b_{ii}(\mathrm{U}) = -\sum_{j \neq i} b_{ij}(\mathrm{U}) = 0\,.$$

If $\alpha_{ij}(\mathrm{U}) = \alpha_{ji}(\mathrm{U})$, then $b_{ij}(\mathrm{U}) = (1-\alpha_{ij}(\mathrm{U}))\, d_{ij}$.

## Nonlinear algebraically stabilized method

$$\sum_{j=1}^{N} (a_{ij} + b_{ij}(\mathrm{U}))\, u_j = f_i\,, \qquad i = 1,\dots,M\,,$$

$$u_i = u_i^b\,, \qquad i = M+1,\dots,N\,.$$

where

$$b_{ij}(\mathrm{U}) = -\max\{(1-\alpha_{ij}(\mathrm{U}))\, a_{ij}, 0, (1-\alpha_{ji}(\mathrm{U}))\, a_{ji}\}\,,$$

$$b_{ii}(\mathrm{U}) = -\sum_{j\neq i} b_{ij}(\mathrm{U}) = 0\,.$$

If $\alpha_{ij}(\mathrm{U}) = \alpha_{ji}(\mathrm{U})$, then $b_{ij}(\mathrm{U}) = (1-\alpha_{ij}(\mathrm{U}))\, d_{ij}$.

If $\min\{a_{ij}, a_{ji}\} \le 0$, then

$$b_{ij}(\mathrm{U}) = \begin{cases} (1-\alpha_{ij}(\mathrm{U}))\, d_{ij} & \text{if } a_{ji} \le a_{ij}\,, \\ (1-\alpha_{ji}(\mathrm{U}))\, d_{ij} & \text{else.} \end{cases}$$

## Monotone Upwind-type Algebraic Stabilization (MUAS)

The above nonlinear algebraically stabilized method with

$$P_i^+ := \sum_{\substack{j \in S_i \\ a_{ij} > 0}} a_{ij}\,(u_i - u_j)^+\,, \qquad P_i^- := \sum_{\substack{j \in S_i \\ a_{ij} > 0}} a_{ij}\,(u_i - u_j)^-\,,$$

$$Q_i^+ := \sum_{j \in S_i} s_{ij}\,(u_j - u_i)^+\,, \qquad Q_i^- := \sum_{j \in S_i} s_{ij}\,(u_j - u_i)^-\,,$$

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

$$\alpha_{ij} := \begin{cases} R_i^+ & \text{if } u_i > u_j\,, \\ 1 & \text{if } u_i = u_j\,, \\ R_i^- & \text{if } u_i < u_j\,. \end{cases} \qquad s_{ij} = \max\{|a_{ij}|, a_{ji}\}$$

DMP guaranteed with arbitrary meshes and data!

Equivalent to the AFC scheme with Kuzmin's limiter if

$\min\{a_{ij}, a_{ji}\} \le 0 \quad \forall\, i \neq j$ and $s_{ij} = |d_{ij}|$.

## Convergence studies for the AFC scheme with Kuzmin's limiter

Many results obtained already in
Barrenechea, John, K., SINUM (2016) for

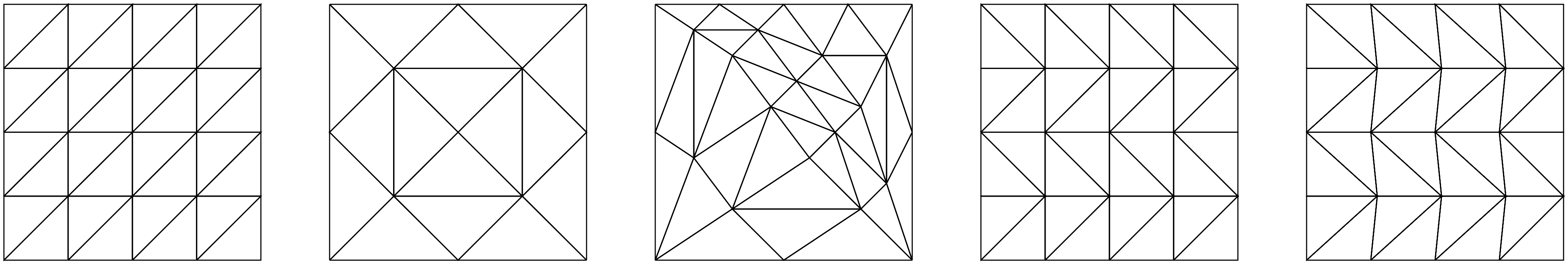
### Example 1

$$\Omega = (0,1)^2, \qquad \varepsilon = 10^{-8}, \qquad \mathbf{b} = (3,2), \qquad c = 1, \qquad u_b = 0.$$

The right-hand side $f$ is chosen such that

$$u(x,y) = 100\,x^2\,(1-x)^2\,y\,(1-y)\,(1-2y)$$

is the exact solution.

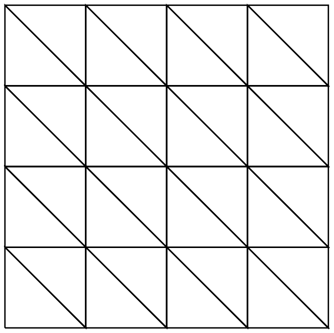
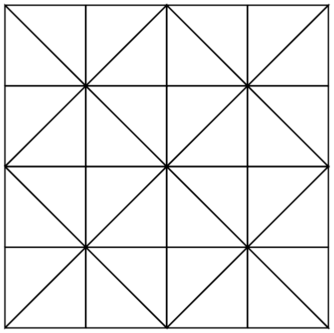
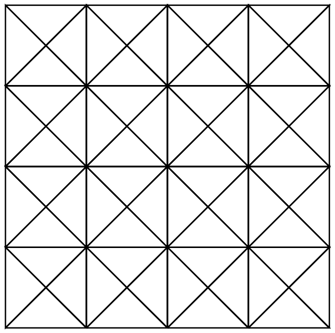
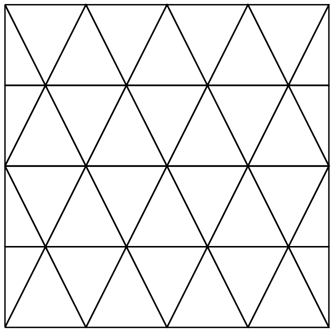
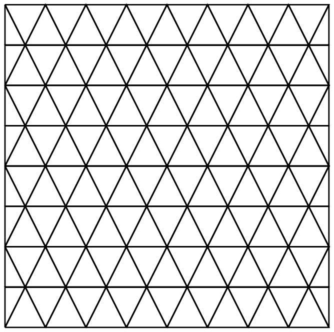


Grids 1 – 5 (left to right)

# Example 1, numerical results for Grid 1

| $ne$ | $\|u-u_h\|_{0,\Omega}$ | ord. | $\|u-u_h\|_{1,\Omega}$ | ord. | $\|u-u_h\|_h$ | ord. |
|---|---|---|---|---|---|---|
| 16 | 1.934e−2 | 1.60 | 4.937e−1 | 0.98 | 5.007e−2 | 1.87 |
| 32 | 5.359e−3 | 1.85 | 2.305e−1 | 1.10 | 1.149e−2 | 2.12 |
| 64 | 1.385e−3 | 1.95 | 1.082e−1 | 1.09 | 2.649e−3 | 2.12 |
| 128 | 3.442e−4 | 2.01 | 5.154e−2 | 1.07 | 6.152e−4 | 2.11 |
| 256 | 8.536e−5 | 2.01 | 2.566e−2 | 1.01 | 1.586e−4 | 1.96 |
| 512 | 2.126e−5 | 2.01 | 1.342e−2 | 0.93 | 3.876e−5 | 2.03 |

analogous results for

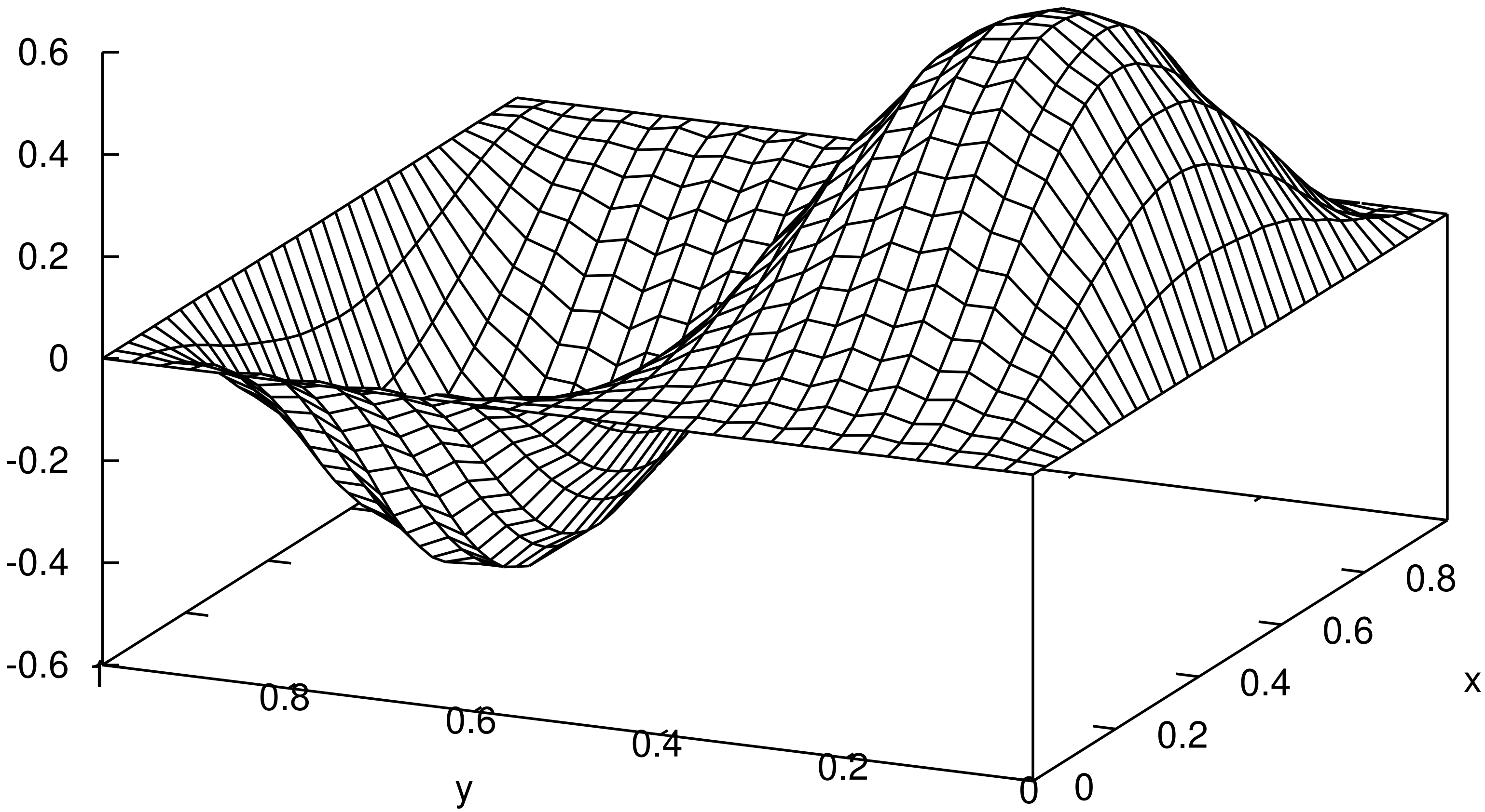






## Example 1, numerical results for Grid 4

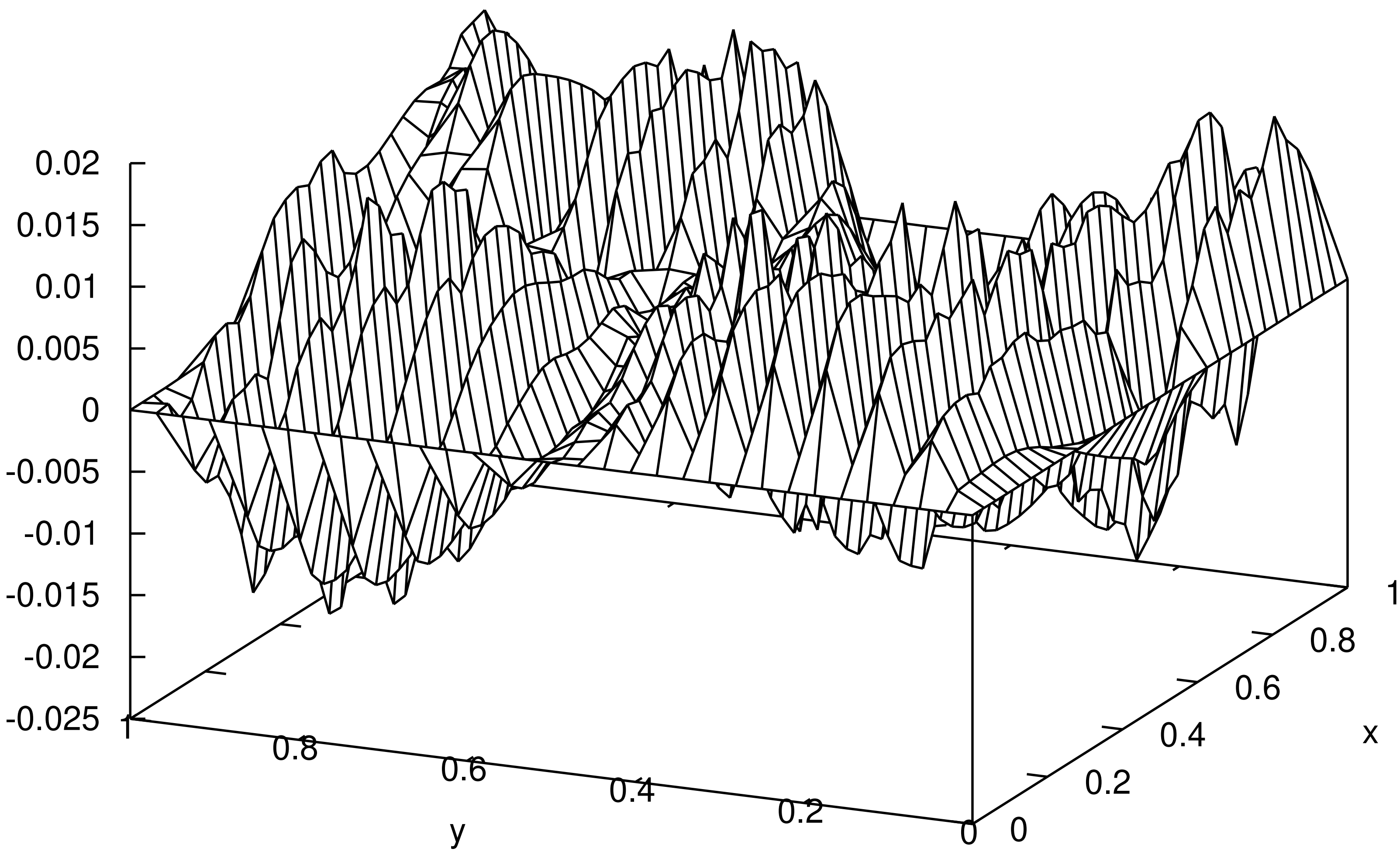
| $ne$ | $\|u-u_h\|_{0,\Omega}$ | ord. | $\lvert u-u_h\rvert_{1,\Omega}$ | ord. | $\|u-u_h\|_h$ | ord. |
|---|---|---|---|---|---|---|
| 16 | 2.019e−2 | 1.65 | 6.005e−1 | 0.68 | 5.663e−2 | 1.74 |
| 32 | 6.285e−3 | 1.68 | 4.832e−1 | 0.31 | 2.138e−2 | 1.41 |
| 64 | 2.308e−3 | 1.45 | 4.549e−1 | 0.09 | 9.485e−3 | 1.17 |
| 128 | 1.092e−3 | 1.08 | 4.442e−1 | 0.03 | 4.490e−3 | 1.08 |
| 256 | 5.543e−4 | 0.98 | 4.368e−1 | 0.02 | 2.187e−3 | 1.04 |
| 512 | 2.823e−4 | 0.97 | 4.327e−1 | 0.01 | 1.083e−3 | 1.01 |

analogous results for Grids 2, 3, 5

# Example 1, approximate solution on Grid 4



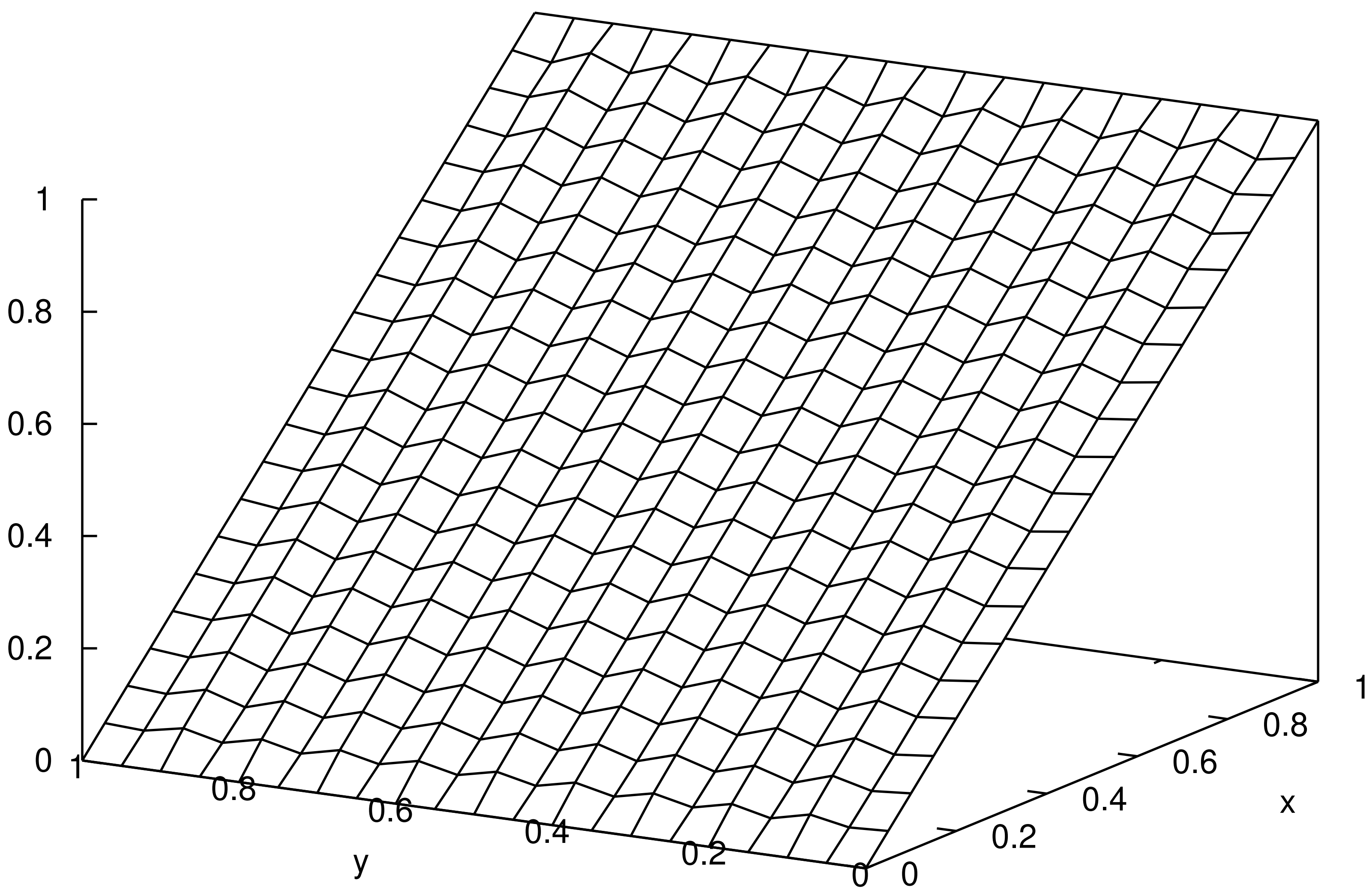
# Example 1, error $u_h - i_h u$ on Grid 4



# Example 2

$\Omega = (0,1)^2$, $\varepsilon = 10^{-8}$, $\mathbf{b} = (1,0)$, $c = 0$, $f = 1$, $u_b(x,y) = x$

# Example 2, approximate solution on Grid 4



# Example 2, numerical results for Grid 4

| $ne$ | $\|u-u_h\|_{0,\Omega}$ | ord. | $\lvert u-u_h\rvert_{1,\Omega}$ | ord. | $\|u-u_h\|_h$ | ord. |
|---|---|---|---|---|---|---|
| 16 | 8.104e−3 | 0.82 | 4.401e−1 | -0.20 | 1.179e−2 | 0.78 |
| 32 | 4.291e−3 | 0.92 | 4.700e−1 | -0.09 | 6.227e−3 | 0.92 |
| 64 | 2.204e−3 | 0.96 | 4.851e−1 | -0.05 | 3.157e−3 | 0.98 |
| 128 | 1.117e−3 | 0.98 | 4.926e−1 | -0.02 | 1.580e−3 | 1.00 |
| 256 | 5.618e−4 | 0.99 | 4.963e−1 | -0.01 | 7.893e−4 | 1.00 |
| 512 | 2.817e−4 | 1.00 | 4.982e−1 | -0.01 | 3.974e−4 | 0.99 |

## Example 2, analytical investigations for Grid 4

the discrete solution $u_h$ is approximately given by

$$u_h(x,y) = x + \alpha \qquad \text{along odd horizontal grid lines},$$

$$u_h(x,y) = x - \beta \qquad \text{along even horizontal grid lines}$$

## Example 2, analytical investigations for Grid 4

the discrete solution $u_h$ is approximately given by

$$u_h(x,y) = x + \alpha \qquad \text{along odd horizontal grid lines},$$

$$u_h(x,y) = x - \beta \qquad \text{along even horizontal grid lines}$$

this function satisfies the Galerkin discretization

up to the perturbation $2\,(\alpha + \beta)\,\varepsilon$

## Example 2, analytical investigations for Grid 4

the discrete solution $u_h$ is approximately given by

$$u_h(x,y) = x + \alpha \qquad \text{along odd horizontal grid lines}\,,$$
$$u_h(x,y) = x - \beta \qquad \text{along even horizontal grid lines}$$

this function satisfies the Galerkin discretization

up to the perturbation $2\,(\alpha+\beta)\,\varepsilon$

$\Rightarrow$ the AFC term should be nearly zero

## Example 2, analytical investigations for Grid 4

the discrete solution $u_h$ is approximately given by

$$u_h(x,y) = x + \alpha \qquad \text{along odd horizontal grid lines}\,,$$
$$u_h(x,y) = x - \beta \qquad \text{along even horizontal grid lines}$$

this function satisfies the Galerkin discretization

up to the perturbation $2\,(\alpha+\beta)\,\varepsilon$

$\Rightarrow$ the AFC term should be nearly zero

Indeed, the AFC term vanishes for $\alpha+\beta = \dfrac{h}{2}\,\dfrac{h}{h-3\,\varepsilon}$.

## Example 2, analytical investigations for Grid 4

Lemma For $u \in P_1(\mathbb{R}^2)$, Example 2, and Grid 4, the quantities from the definition of the Kuzmin limiter satisfy

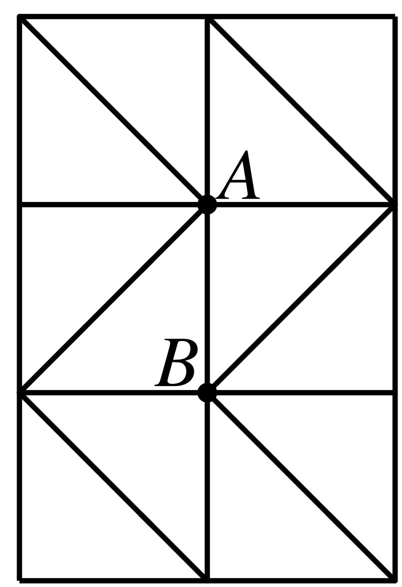
$$P_A^+ \le Q_A^+\,, \qquad P_A^- \ge Q_A^-$$

and

$$P_B^+ \le \frac{2h-3\varepsilon}{h-3\varepsilon}\,Q_B^+\,, \qquad P_B^- \ge \frac{2h-3\varepsilon}{h-3\varepsilon}\,Q_B^-\,,$$

where the latter inequalities are sharp

$\Rightarrow$ the AFC scheme is not linearity preserving.



## Example 2, analytical investigations for Grid 4

Lemma For $u \in P_1(\mathbb{R}^2)$, Example 2, and Grid 4, the quantities from the definition of the Kuzmin limiter satisfy

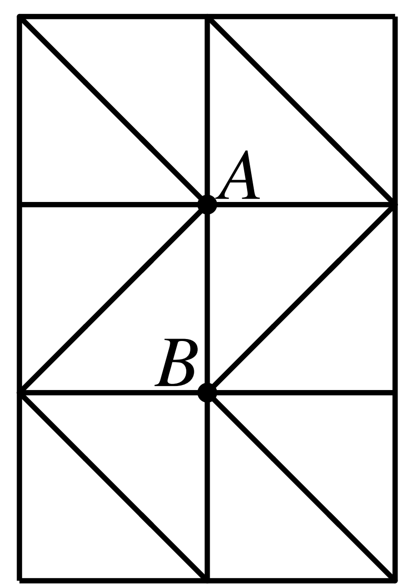
$$P_A^+ \le Q_A^+\,, \qquad P_A^- \ge Q_A^-$$

and

$$P_B^+ \le \frac{2\,h - 3\,\varepsilon}{h - 3\,\varepsilon}\,Q_B^+\,, \qquad P_B^- \ge \frac{2\,h - 3\,\varepsilon}{h - 3\,\varepsilon}\,Q_B^-\,,$$

where the latter inequalities are sharp

$\Rightarrow$ the AFC scheme is not linearity preserving.




Possible remedy:

$$R_B^+ = \min\left\{1, \frac{\mu\,Q_B^+}{P_B^+}\right\}, \quad R_B^- = \min\left\{1, \frac{\mu\,Q_B^-}{P_B^-}\right\}$$

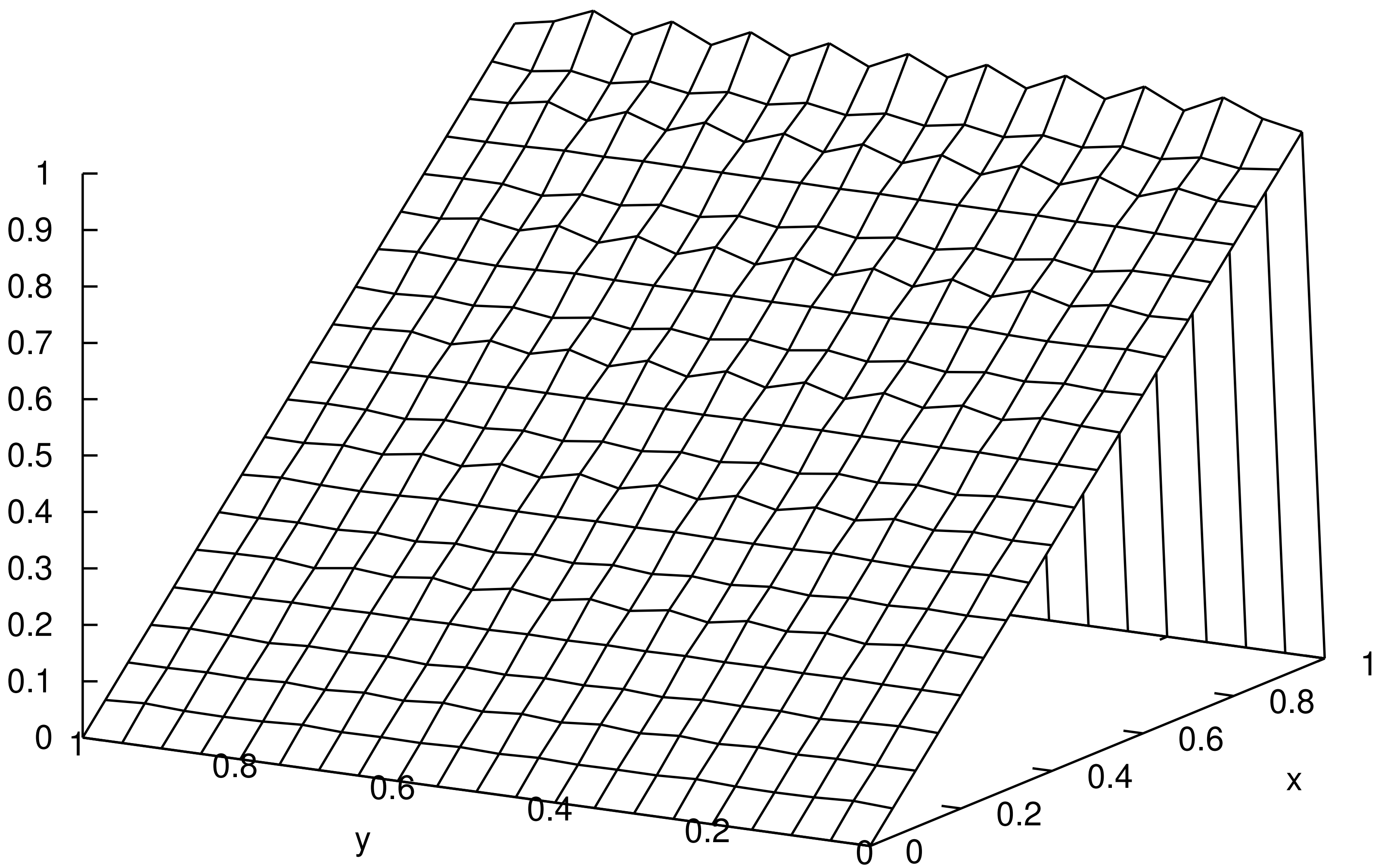
with $\mu = (2\,h - 3\,\varepsilon)/(h - 4\,\varepsilon)$.

## Example 3

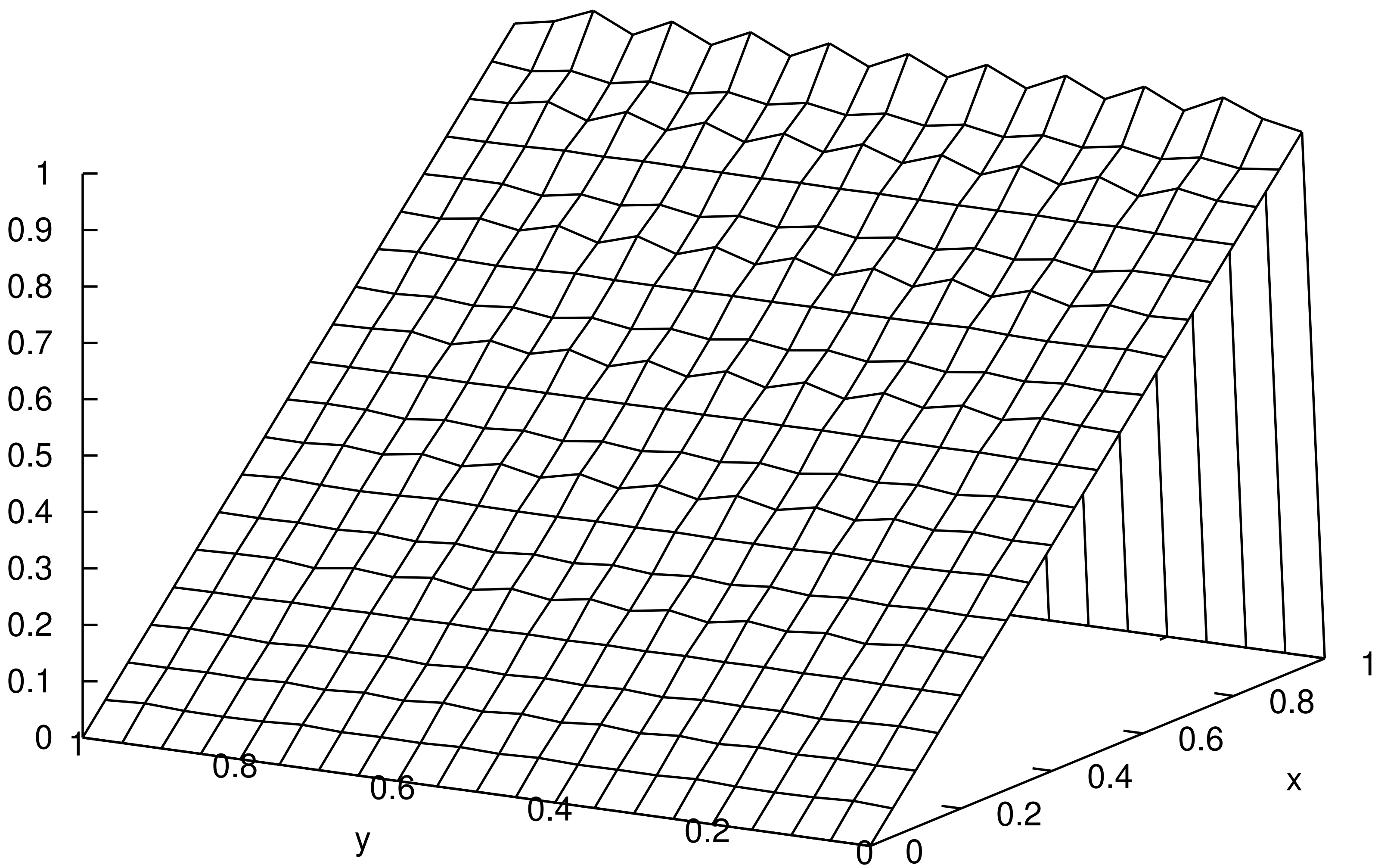
$\Omega = (0,1)^2$, $\varepsilon = 10^{-8}$, $\mathbf{b} = (1,0)$, $c = 0$, $f = 1$,

$$u_b(x,y) = x - \frac{\mathrm{e}^{\frac{x}{\varepsilon}} - 1}{\mathrm{e}^{\frac{1}{\varepsilon}} - 1}\,.$$

# Example 3, approximate solution on Grid 4 with modified $R_B^{\pm}$

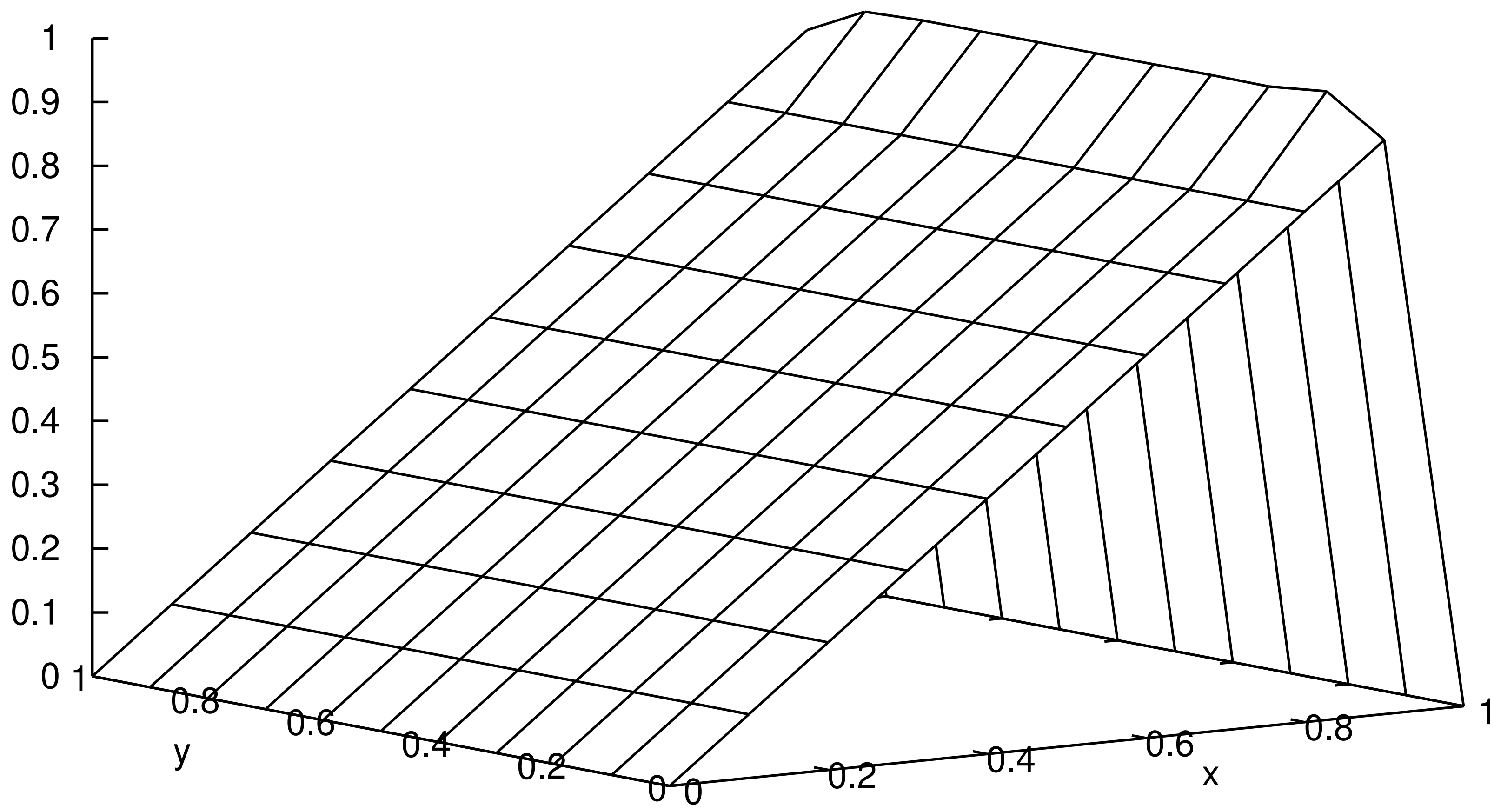


## Example 3, approximate solution on Grid 4 with modified $R_B^{\pm}$

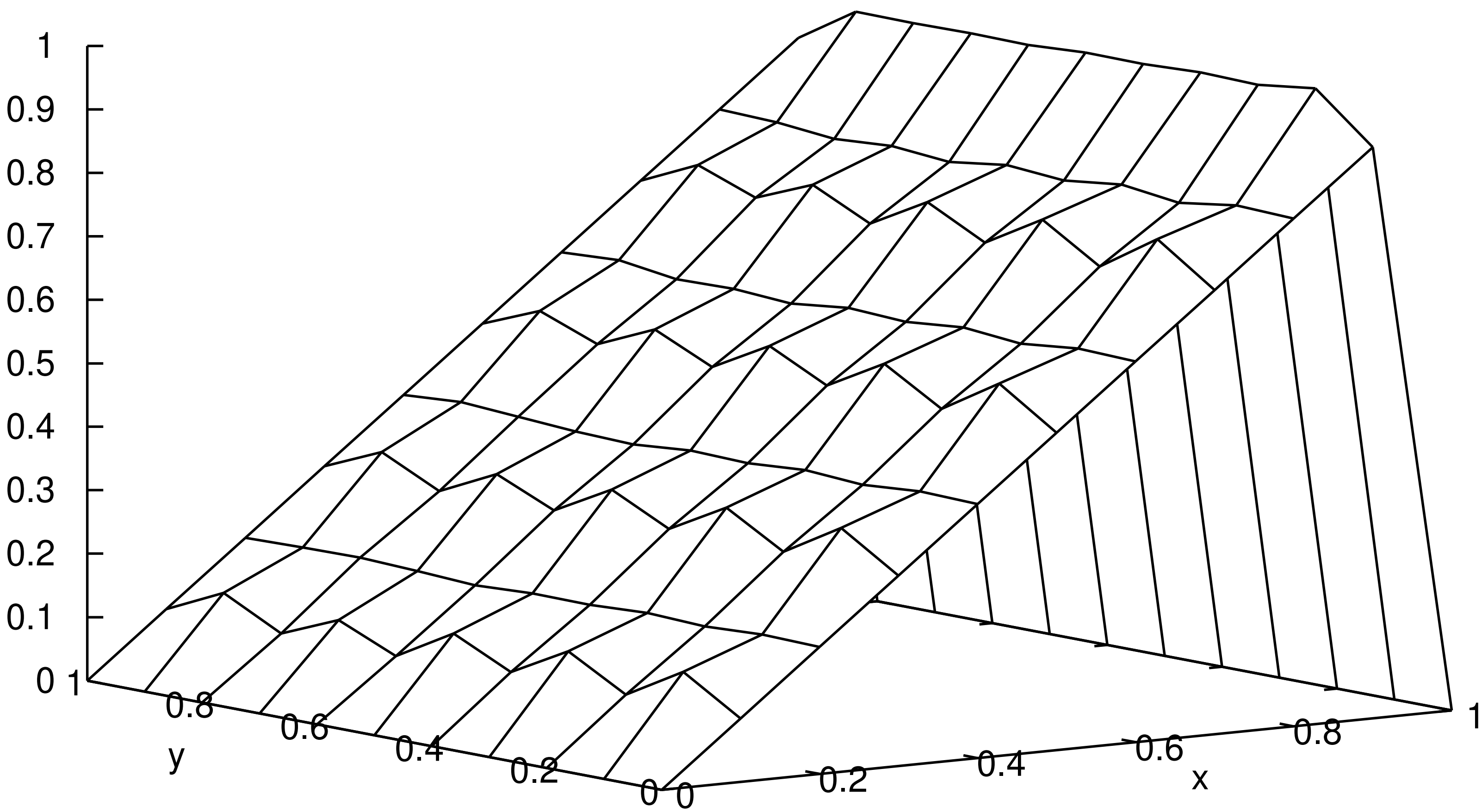



It is possible to show analytically that the AFC term again vanishes for oscillating functions of this type!

# Example 3, approximate solution on Grid 1



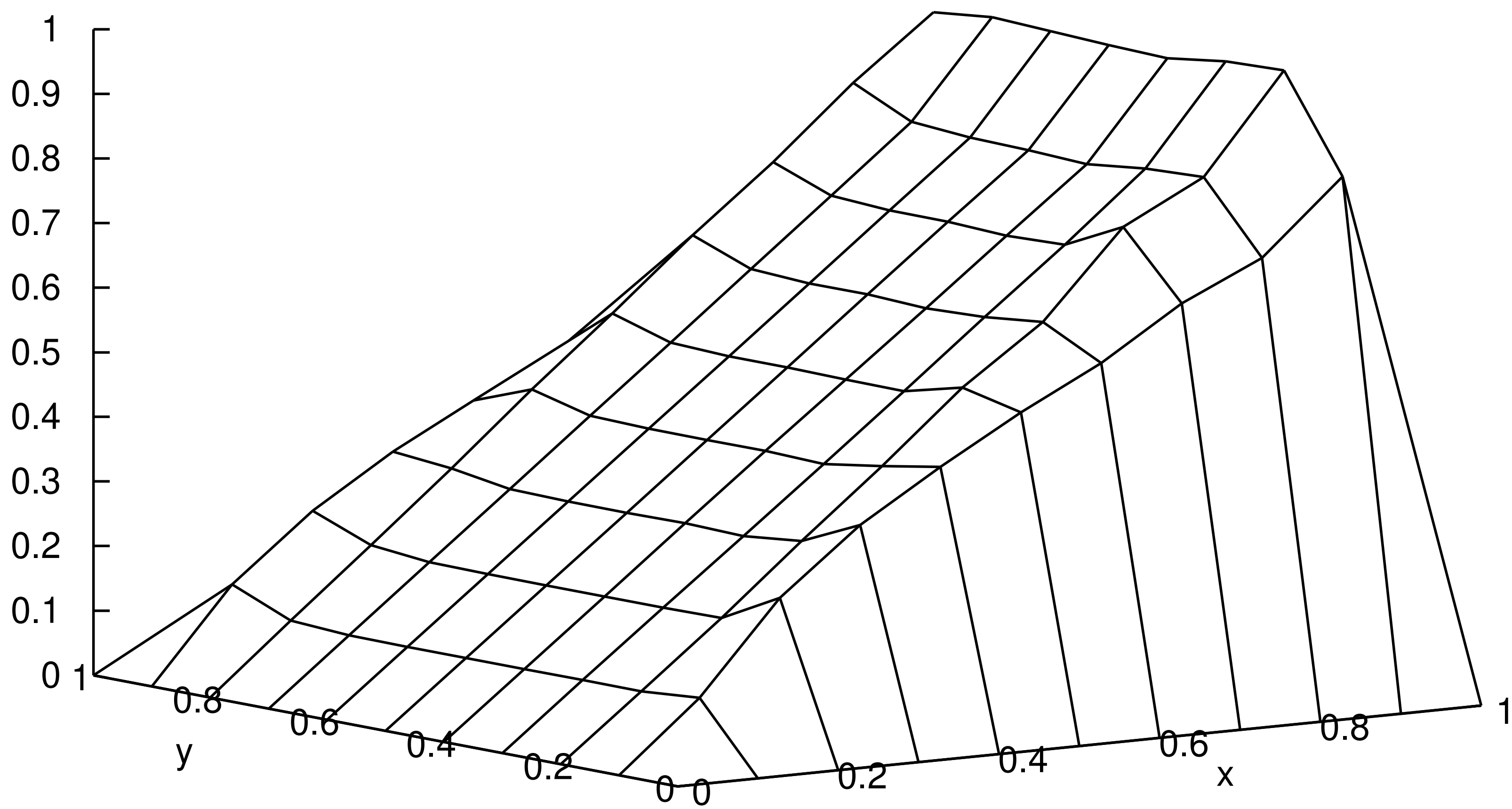
# Example 3, approximate solution on Grid 1 with $R_i^{\pm}$ modified using $\mu = 2$



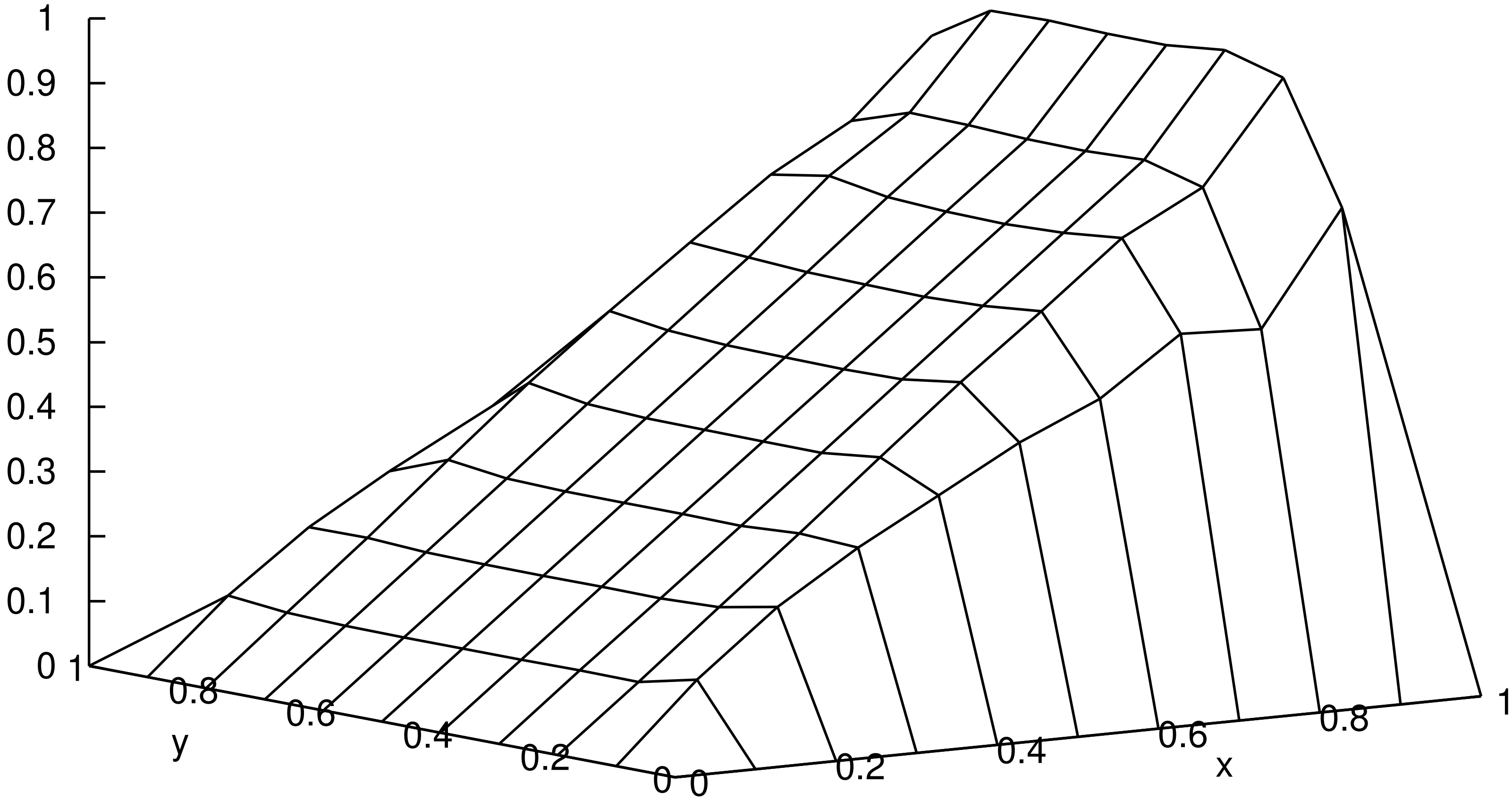
# Example 4

$\Omega = (0,1)^2$, $\varepsilon = 10^{-8}$, $\mathbf{b} = (1,0)$, $c = 0$, $f = 1$, $u_b = 0$

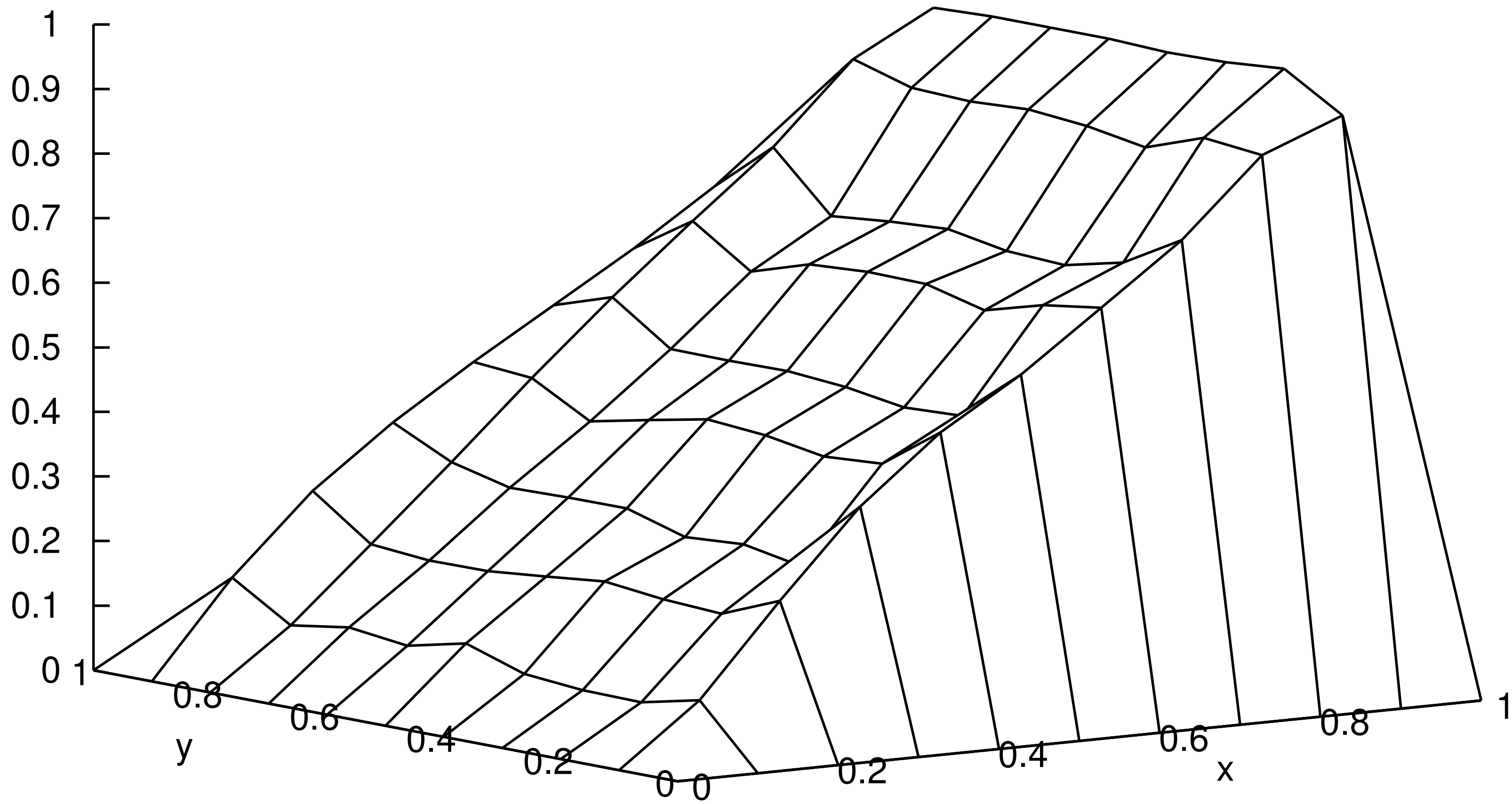
# Example 4, approximate solution on Grid 1



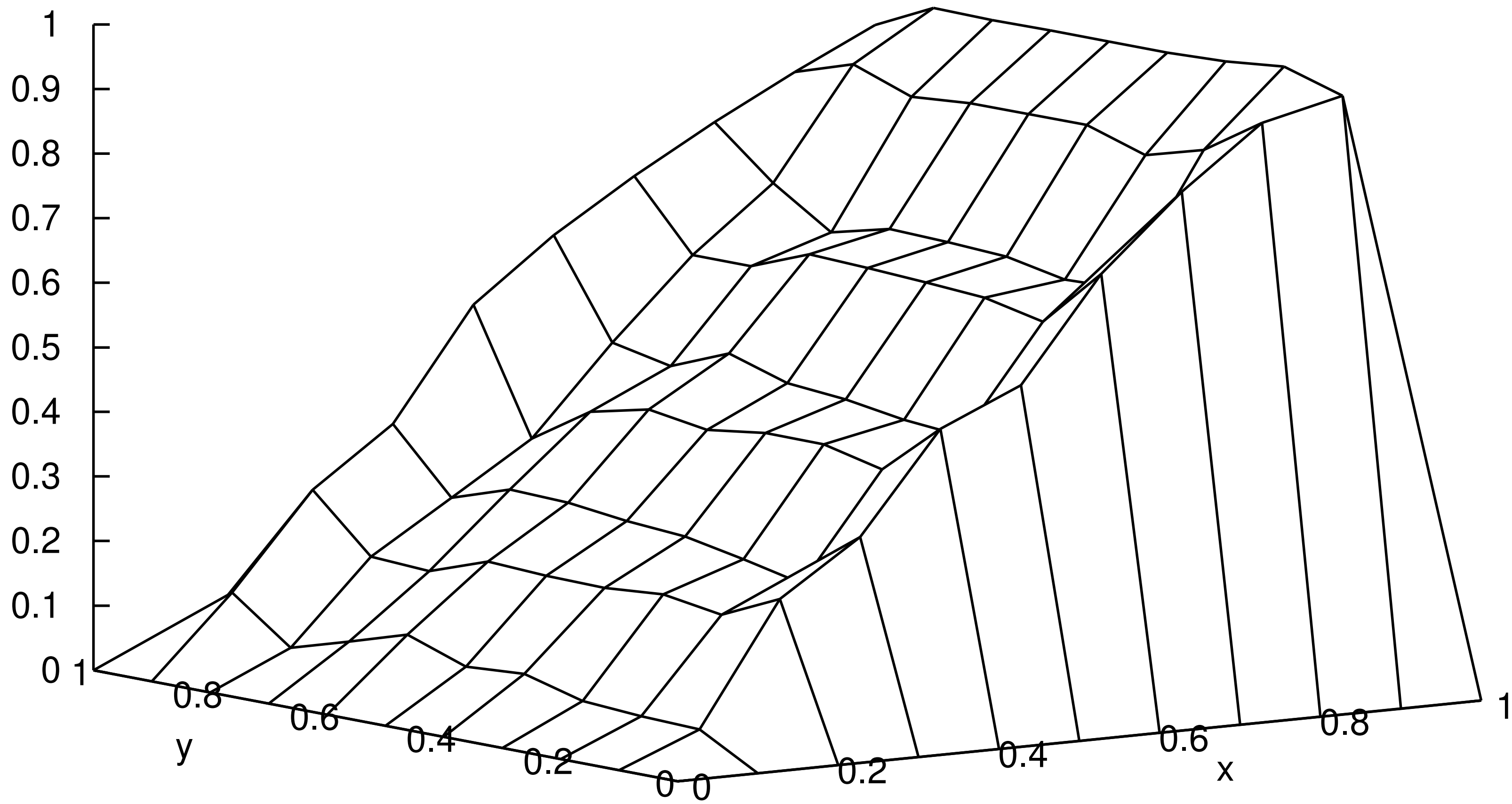
# Example 4, approximate solution on Grid 1, $P_i^{\pm}$ as in the BJK limiter



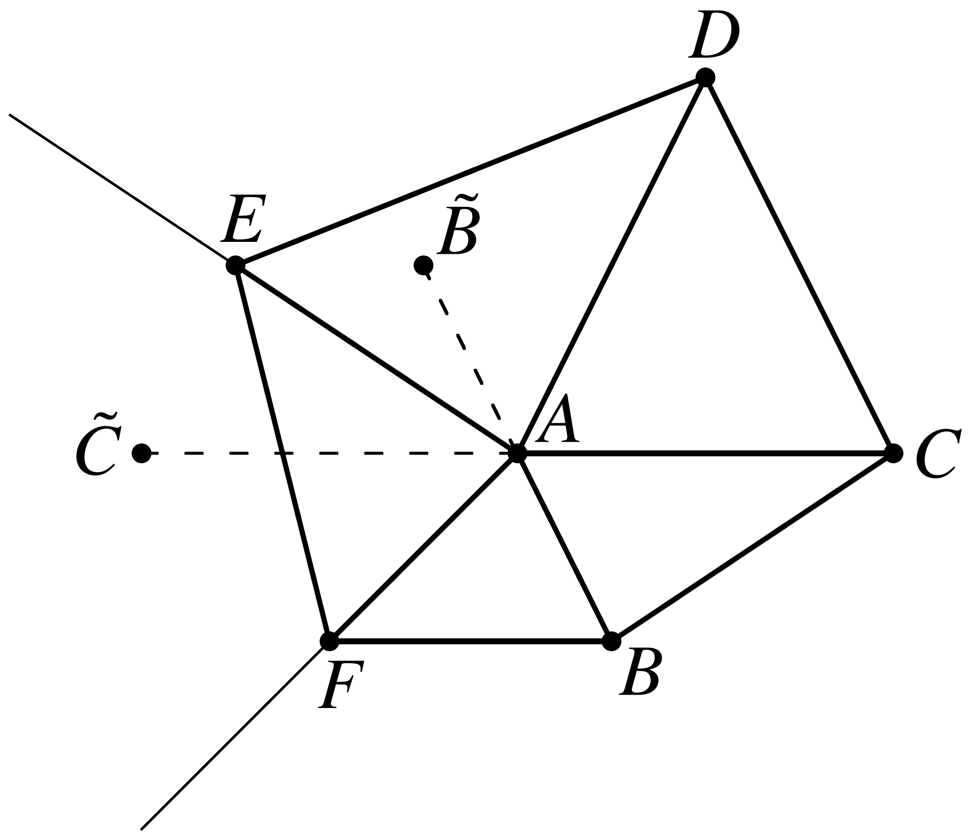
# Example 4, approximate solution on Grid 1 for the BJK limiter with $\mu_i = 1$



## Example 4, approximate solution on Grid 1 for the BJK limiter with $\mu_i = 2$



# Symmetrization of the MUAS method




$$u_{AC} = u_A + \nabla u_h|_{AEF} \cdot (\tilde{C} - A) = u_A + \nabla u_h|_{AEF} \cdot (A - C)\,.$$

## Symmetrized Monotone Upwind-type Algebraic Stabilization (SMUAS)

The quantities $P_i^{\pm}$, $Q_i^{\pm}$ in the definition of the MUAS method are replaced by

$$P_i^+ = \sum_{j\in S_i} |d_{ij}| \left\{ (u_i - u_j)^+ + (u_i - u_{ij})^+ \right\},$$

$$P_i^- = \sum_{j\in S_i} |d_{ij}| \left\{ (u_i - u_j)^- + (u_i - u_{ij})^- \right\},$$

$$Q_i^+ = \sum_{j\in S_i} s_{ij} \left\{ (u_j - u_i)^+ + (u_{ij} - u_i)^+ \right\},$$

$$Q_i^- = \sum_{j\in S_i} s_{ij} \left\{ (u_j - u_i)^- + (u_{ij} - u_i)^- \right\}$$

with

$$u_{ij} = u_i + \nabla u_h|_{T_{ij}} \cdot (x_i - x_j)$$

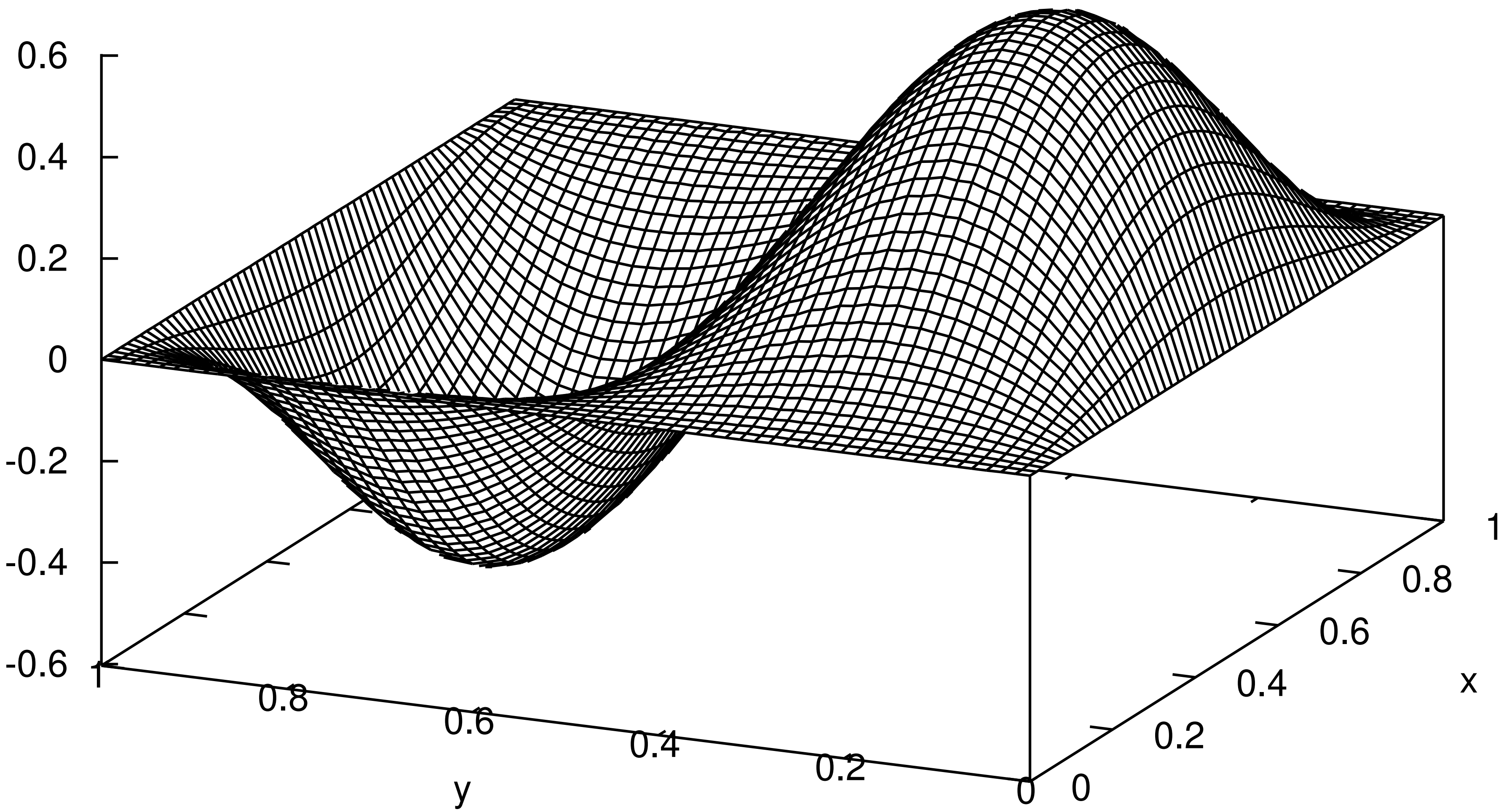
## Symmetrized Monotone Upwind-type Algebraic Stabilization (SMUAS)

The SMUAS method is linearity preserving and satisfies Assumptions (A1) and (A2) on arbitrary meshes. For regular families of triangulations, one also has the above error estimate.

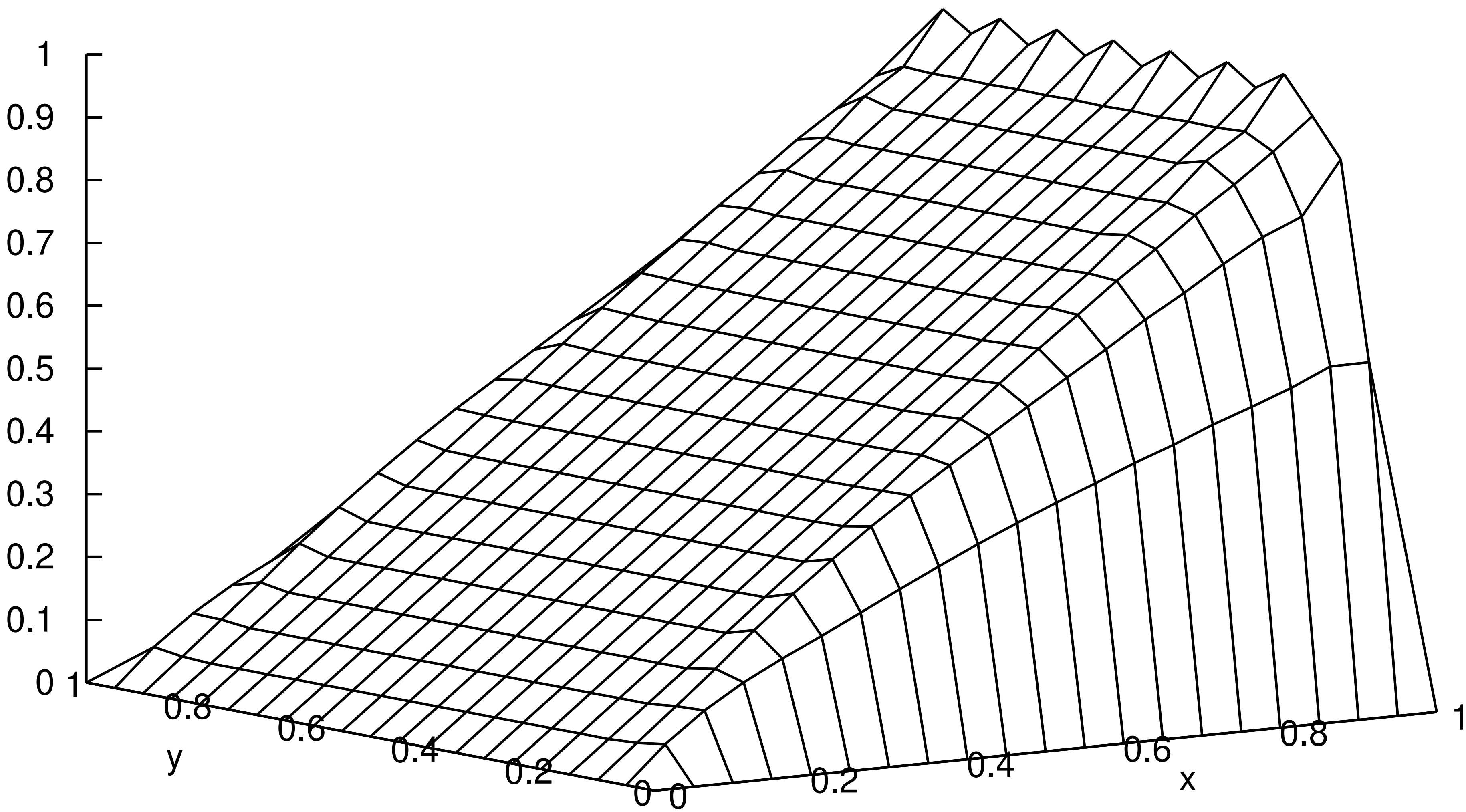
## Example 1, numerical results for the SMUAS method on Grid 4

| $ne$ | $\|u-u_h\|_{0,\Omega}$ | ord. | $\|u-u_h\|_{1,\Omega}$ | ord. | $\|u-u_h\|_h$ | ord. |
|---|---|---|---|---|---|---|
| 16 | 2.147e−2 | 1.61 | 4.734e−1 | 0.98 | 5.530e−2 | 1.92 |
| 32 | 6.353e−3 | 1.76 | 2.529e−1 | 0.90 | 1.479e−2 | 1.90 |
| 64 | 1.783e−3 | 1.83 | 1.363e−1 | 0.89 | 3.922e−3 | 1.92 |
| 128 | 4.706e−4 | 1.92 | 7.220e−2 | 0.92 | 1.054e−3 | 1.90 |
| 256 | 1.221e−4 | 1.95 | 3.807e−2 | 0.92 | 2.940e−4 | 1.84 |
| 512 | 3.135e−5 | 1.96 | 2.002e−2 | 0.93 | 7.896e−5 | 1.90 |

# Example 1, solution of the SMUAS method on Grid 4



# Example 4, solution of the SMUAS method on Grid 4



# Conclusions

– the SMUAS method seems to have all the properties we wish

BUT

– a theoretical explanation of the observed optimal convergence behaviour is still open