Application of Multigrid to an Initial Control Problem

Michael Czajkowski

Universität Düsseldorf

Mathematisches Institut

Lehrstuhl für Angewandte Mathematik

Problem formulation

Initial control problem (icp) for the wave equation considered here.

Underlying direct problem:
compute the solution $u = Mu_0$ of the wave equation

$$\begin{aligned} u_{tt} - u_{xx} - u_{yy} &= f && \text{on} \quad \Omega \times (0,T) \\ u(x,y,t) &= 0 && \text{on} \quad \partial\Omega \times [0,T] \\ u(x,y,t=0) &= u_0 && \text{on} \quad \Omega \\ u_t(x,y,t=0) &= 0 && \text{on} \quad \Omega \end{aligned}$$

for a given initial function u_0.

Formulation of the initial control problem as a special inverse problem:

given: finitely many data $d_j \in \mathbb{R}$ at discrete points (x_j, y_j, t_j), $j = 1, \ldots, j_{\max}$

now: find an initial function u_0 satisfying (ideal case)

$$u(x_j, y_j, t_j) = (Mu_0)(x_j, y_j, t_j) = d_j, \quad j = 1, \ldots, j_{\max}$$

Challenging time-dependent inverse problems in practice: weather prediction, geophysics.

Problem:
icp will almost always have no unique solution u_0 continuously depending on the data ("ill-posedness of inverse problems")

Remedy:

– Solution of the least squares problem with minimal norm ("generalized inverse"; theoretically)
– Regularization methods (due to Tikhonov 1963; practically)

Starting point for such regularization methods:

$$\boxed{\begin{array}{c}\text{minimization functional} = \\ \text{regularization term} + \text{least squares term}\end{array}}$$

Minimization functional for the icp considered here:

$$J_h^\alpha : H_h \longrightarrow \mathbb{R}$$

$$J_h^\alpha(u_h) = \alpha \cdot \int\limits_\Omega \left(\frac{\partial}{\partial x} u_h\right)^2 + \left(\frac{\partial}{\partial y} u_h\right)^2 d\Omega + \frac{1}{j_{\max}} \sum_{j=1}^{j_{\max}} (M_h(x_j, y_j, t_j) u_h - d_j)^2, \ \alpha > 0$$

now determine minimal solution

$$u_h^*(\alpha) \in H_h : \qquad J_h^\alpha(u_h^*(\alpha)) = \min_{u_h \in H_h} J_h^\alpha(u_h)$$

Properties of the minimization functional and the minimal solutions $u_h^*(\alpha)$

- Essential result: the adjoint operator in the corresponding linear equation can be interpreted as the time-reverse solution of a similar wave equation.
- For all $\alpha > 0$ there is a unique minimal solution $u_h^*(\alpha)$ of J_h^α.
- For $\dim H_h > j_{\max}$ the pure least squares problem has no unique solution.
- For two complete contiguous time layers as data the pure least squares problem is regular.
- Error estimate for a regular least squares problem:

  $$\|u_h^*(0) - u_0\|_\infty \le Ch^2 + K(h) \cdot f(h)$$

  where $f(h)$ denotes the convergence estimate for the discretization of the wave equation.
- Convergence property for fixed space H_h (also for a singular least squares problem):

  $$\lim_{\alpha \to 0} u_h^*(\alpha) = u_h^*(0)$$

- Boundedness of the minimal solutions $u_h^*(\alpha)$ in the H^1-norm for fixed regularization parameter:

  $$\forall H_h : \quad \|u_h^*(\alpha)\|_{H^1} < \infty$$

Numerical computation of the minimal solutions $u_h^*(\alpha)$

Only defect-oriented iterative solvers possible for efficiency reasons

Typical methods:

- Richardson iteration / gradient method as the most simple procedure, also taken as smoothing iteration in multigrid:

$$x^{i+1} = x^i + \beta \cdot (b - Ax^i) \quad (Ax = b)$$

- cg iteration as a nearly optimal method for general positive definite problems
- multigrid as a highly efficient procedure with respect to the regularization term (discrete laplacian), starting from a variational formulation (due to Mandel, McCormick et al., 1982)

Theoretical analysis of the iterative solvers

- Same computational amount for cg and gradient method, only slightly higher computational amount for multigrid
- Convergence estimate for cg and gradient method by the spectral condition number
$$\kappa = \mathcal{O}\left(\frac{n^2}{\alpha}\right) \qquad (n^2 = \dim H_h)$$
- An exact analysis of the properties of Richardson iteration and multigrid can be carried out in the case of complete and contiguous time layers as data by means of discrete Fourier analysis:
  - explicit formulae for the optimal parameter in Richardson iteration and the spectral radius
  - determination of optimal regularization parameters α with respect to Richardson iteration convergence.
  - explicit formulae for smoothing and convergence rates in multigrid
  - analysis and replacement of inefficient Galerkin coarse-grid operators in multigrid

Multigrid (1)
Formulation in a variational framework
(McCormick 1982)

Let $H_0 < H_1 < \ldots < H_m$ be nested finite-dimensional Hilbert spaces with scalar products $\langle \cdot , \cdot \rangle_k$, $k = 0, 1, \ldots, m$, respectively, and canonical norms $\| \cdot \|_k$. Let $a(\cdot , \cdot)$ be a s. p. d. bilinear form for H_m and $\| \cdot \|_E$ the norm induced by a. For $f_m \in H_m$ we consider the variational equation

$$u_m^* \in H_m : \quad \forall\, v_m \in H_m : \quad a(u_m^*, v_m) = \langle f_m, v_m \rangle_m$$

and for $f_k \in H_k$, $k = 0, \ldots, m-1$, similar equations

$$u_k^* \in H_k : \quad \forall\, v_k \in H_k : \quad a(u_k^*, v_k) = \langle f_k, v_k \rangle_k .$$

$F_k, G_k : H_k \times H_k \to H_k$ denote smoothing operators, i. e., operators with the consistence property

$$F_k(u_k^*, f_k) = u_k^* = G_k(u_k^*, f_k).$$

The following algorithm then describes a multigrid V-cycle ($\mu = 1$) and a W-cycle ($\mu = 2$), respectively.

Algorithm:

1) If $k = 0$, then $u_0 := G_0(F_0(u_0, f_0), f_0)$
   or exact solution: $u_0 := u_0^*$

2) If $k > 0$, then

   a) $u_k := F_k(u_k, f_k)$ (*pre-smoothing step*)

   b) perform μ iterations of this algorithm on H_{k-1} with fixed value of u_k applied to the following correction problem starting from the initial value $u_{k-1} = 0$:

$$A_{k-1}u_{k-1}^* = f_{k-1} \quad \text{with}$$
$$f_{k-1} = I_k^{k-1}(f_k - A_k u_k), \; A_{k-1} = I_k^{k-1} A_k I_{k-1}^k$$

   c) $u_k := u_k + u_{k-1}$ (*correction*)

   d) $u_k := G_k(u_k, f_k)$ (*post-smoothing step*)

Here, the operators $I_{k-1}^k, \; I_k^{k-1}$ and A_k are defined by

$$\forall\, v_k, w_k \in H_k : \qquad \langle A_k v_k, w_k \rangle_k = a(v_k, w_k)$$

$$\forall\, v_k \in H_k \; \forall\, w_{k-1} \in H_{k-1} : \quad \langle I_k^{k-1} v_k, w_{k-1} \rangle_{k-1} = \langle v_k, w_{k-1} \rangle_k$$

$$\forall\, v_{k-1} \in H_{k-1} : \qquad I_{k-1}^k v_{k-1} = v_{k-1},$$
$$\text{i.e.,} \;\; I_{k-1}^k = id_{|H_{k-1}}$$

Multigrid (2)
Convergence analysis

Representation for the smoothing operators F_k, G_k:

$$u_k^{(i+1)} = M_k u_k^{(i)} + s_k, \quad i = 0, 1, \ldots$$

with linear operators $M_k \in \mathcal{L}(H_k, H_k)$, $k = 0, \ldots, m$

Fundamental convergence result:

Theorem 1: *Let the operators $M_k \in \mathcal{L}(H_k, H_k)$ be contracting in the energy norm, i. e., for all $k = 0, \ldots, m$*

$$\|M_k\|_E \leq \xi < 1.$$

Then, we have for the multigrid operator M_{MG} with $\mu = 1$

$$\|M_{\mathrm{MG}}\|_E \leq \xi^{\nu_1 + \nu_2},$$

i. e., the reduction of the error in the energy norm by multigrid iteration is at least as good as $\nu_1 + \nu_2$ iterations of the smoothing procedure on H_m.

Sharper convergence results for the two-grid iteration under additional assumptions on the smoothing operators:

Theorem 2: *With the bounds for the linear smoothing operators*

$$\|M_1\|_E \leq \xi < 1, \quad \|M_{1|_{H_0^{\perp_E}}}\|_E \leq \eta \leq \xi < 1$$

and

$$\|M_0\|_E \leq \vartheta < 1$$

we have for the contraction number of the two-grid operator $M_{2\mathrm{G}}$ *performing* ν_c *iterations on the coarse space* H_0

$$\|M_{2\mathrm{G}}\|_E \leq \sqrt{2}\,\xi^{\nu_1} \cdot \max\{\eta^{\nu_2},\, \xi^{\nu_2}\vartheta^{\nu_c}\}.$$

If the correction equation on the coarse space H_0 *is solved exactly, the better estimate holds*

$$\|M_{2\mathrm{G}}\|_E \leq \eta^{\nu_2} \cdot \xi^{\nu_1}.$$

Theorem 3: *We consider the two-grid iteration with exact solution of the correction equation on the coarse space* H_0*. For the linear smoothing operator* $M_1 \in \mathcal{L}(H_1, H_1)$ *we assume*

$$\|M_{1|_{H_0^{\perp_E}}}\|_E \leq \eta < 1, \quad \mathcal{R}(M_{1|_{H_0}}) \subset H_0.$$

Then, for the contraction number

$$\|M_{2G}\|_E \leq \eta^{\nu_1+\nu_2}$$

holds.

Theorem 3 directly applies to the icp with complete time layers as data.

Computation of the optimal parameters, smoothing rate and convergence rate of Richardson iteration and two-grid iteration, respectively, by means of discrete Fourier analysis and comparison with the experimental convergence rate of two-grid iteration in the tables below.

Table 1:
Discretization parameters $n = 80$, $\delta t = 0.002$, one complete time layer at $t = 100 \cdot \delta t$ as data, only one iteration in the pre-smoothing step of two-grid iteration with exact solution on the coarse grid.

α	β_{2G}^{opt}	$\sigma(\beta_{2G}^{\text{opt}})$	β_{R}^{opt}	$\varrho(M_R^{\beta_{\text{opt}}})$	ϱ_{2G}
0.1	0.00031	0.599	0.0004	0.999	0.57
0.01	0.00312	0.598	0.0039	0.998	0.56
0.001	0.0309	0.582	0.039	0.997	0.54
0.0001	0.274	0.618	0.337	0.988	0.59
0.00001	1.178	0.835	1.278	0.991	0.81
0.000001	1.755	0.975	1.775	0.997	0.97
0.0	1.857	$> 1-10^{-5}$	1.857	$> 1-10^{-7}$	>0.99

Table 2:
Discretization parameters $n = 80$, $\delta t = 0.002$, two complete and contiguous time layers at $t = 100 \cdot \delta t$, $101 \cdot \delta t$ as data, only one iteration in the pre-smoothing step of two-grid iteration with exact solution on the coarse grid.

α	$\beta_{2\mathrm{G}}^{\mathrm{opt}}$	$\sigma(\beta_{2\mathrm{G}}^{\mathrm{opt}})$	$\beta_{\mathrm{R}}^{\mathrm{opt}}$	$\varrho(M_{\mathrm{R}}^{\beta_{\mathrm{opt}}})$	$\varrho_{2\mathrm{G}}$
0.1	0.00031	0.599	0.00039	0.999	0.57
0.01	0.00312	0.598	0.0039	0.998	0.57
0.001	0.0309	0.585	0.0390	0.997	0.56
0.0001	0.278	0.612	0.343	0.989	0.58
0.00001	1.209	0.816	1.323	0.987	0.77
0.000001	1.816	0.949	1.857	0.993	0.91
0.0	1.923	0.972	1.946	0.995	0.96

Multigrid (3) Replacement of the full Galerkin coarse-grid operators

Until now full Galerkin operators used on all grids, i. e., the evaluation of the operator takes the same computational work on all grids: inefficient multigrid.

Theoretical analysis of replacement:

Theorem 4: *We consider the two-grid iteration. For the linear smoothing operator* $M_1 \in \mathcal{L}(H_1, H_1)$ *we assume*

$$\|M_1\|_E \leq \xi < 1, \quad \|M_{1|_{H_0^{\perp_E}}}\|_E \leq \eta \leq \xi < 1.$$

The coarse-grid correction equation with the full Galerkin coarse-grid operator

$$A_0 u_0^* = f_0, \;\; A_0 = I_1^0 A_1 I_0^1$$

will be replaced by

$$\tilde{A}_0 \tilde{u}_0^* = f_0$$

with an invertible operator $\tilde{A}_0 \in \mathcal{L}(H_0, H_0)$. *For* $\varepsilon \in \mathbb{R}_0^+$ *let the estimate*

$$\|\tilde{A}_0^{-1}(\tilde{A}_0 - A_0)\|_E \leq \varepsilon$$

hold. Then, for the contraction number of the two-grid operator $\tilde{M}_{2\mathrm{G}}$ *with exact solution of the modified correction equation we obtain*

$$\|\tilde{M}_{2\mathrm{G}}\|_E \leq \sqrt{2}\,\xi^{\nu_1} \cdot \max\,\{\eta^{\nu_2}, \xi^{\nu_2}\varepsilon\}.$$

A replacement with small ε should be uncritical.

Efficient coarse grid operators: representation operators for similar minimization functionals on the coarser grids.

Again, a precise analysis of $\|\tilde{A}_0^{-1}(\tilde{A}_0 - A_0)\|_E$ can be carried out in the case of complete time layers as data.

Table 3: $\|\tilde{A}_0^{-1}(\tilde{A}_0 - A_0)\|_E$ for one complete time layer of data at $t = 0.20$, $n = 20$, $\delta t = 0.01$.

	time step δt_H		
α	0.01	0.02	0.04
0.1	0.49	0.49	0.49
0.01	0.51	0.51	0.51
0.001	0.64	0.62	0.67
0.0001	5.23	5.10	3.66
0.00001	45.80	38.85	39.78

Table 4: $\|\tilde{A}_0^{-1}(\tilde{A}_0 - A_0)\|_E$ for two complete time layers of data at $t = 0.20, 0.21$, $n = 20$, $\delta t = 0.01$.

	time step δt_H		
α	0.01	0.02	0.04
0.1	0.49	0.49	0.49
0.01	0.52	0.52	0.54
0.001	0.68	0.69	0.74
0.0001	4.00	2.39	0.91
0.00001	10.87	4.29	0.95

Numerical examples (1)

Comparison of gradient method, cg and multigrid

Initial function for simulation:

$$u_0(x, y) = \sin 3\pi x \sin 2\pi y.$$

icp: $j_{\max} = 400$ randomly distributed exact data between $t_1 = 0.05$ and $t_2 = 0.2$; discretization parameters $n = 40$ and $\delta t = 0.005$.

3-grid method with V-cycle:

– efficient coarse-grid operators with halving the time stepsize per grid level

– 2 pre- and 3 post-smoothing iterations of the gradient method and 20 relaxation steps on the coarsest grid

– use of coarse-grid correction parameters

– canonical interpolation and restriction with respect to the linear finite element space on the unit square and the resulting 5-point star for the discretization of the regularization term

Table 5: Asymptotic convergence rates of gradient method and multi-grid

procedure	regularization parameter α				
	0.000001	0.00001	0.00003	0.0001	0.0003
gradient	0.9955	0.97	0.93	0.87	0.82
3-grid	0.9902	0.91	0.80	0.53	0.32

Table 6: Relative maximum error in percent between the iterates and the exact initial function

procedure	#It.	regularization parameter α			
		0.000001	0.00001	0.00003	0.0001
∞		15.1	15.8	16.8	19.8
gradient	5	136.7	111.3	84.4	51.6
	10	131.1	93.6	56.1	29.7
	50	109.1	31.0	17.7	19.8
cg	5	137.1	111.4	88.0	56.7
	10	120.5	98.2	37.4	25.4
	50	69.2	15.8	16.8	19.8
3-grid	5	128.4	86.1	49.2	23.6
	10	119.5	60.0	25.0	19.9
	50	82.4	16.8	16.8	19.8

Numerical examples (2)

Comparison of cg and multigrid for perturbed data by some pictures

- Initial function for simulation:

$$u_0(x,y) = \sin 2\pi x\ \sin \pi y\ +\ \frac{1}{4}\sin 4\pi x\ \sin 3\pi y.$$

- icp: $j_{\max} = 1000$ randomly distributed perturbed data between $t_1 = 0.05$ and $t_2 = 0.2$, maximum perturbation of 10%; discretization parameters $n = 80$ and $\delta t = 0.0025$.
- 4-grid method with V-cycle and same parameters as in the last example

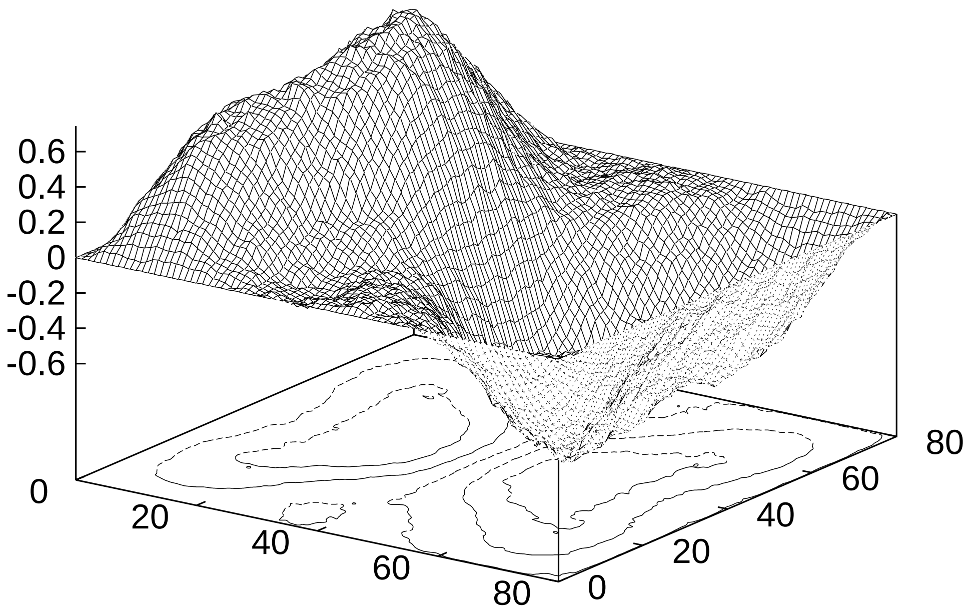

Figure 1.1: Exact discrete solution of the icp, maximum data perturbation 10%, regularization parameter $\alpha = 0.000003$

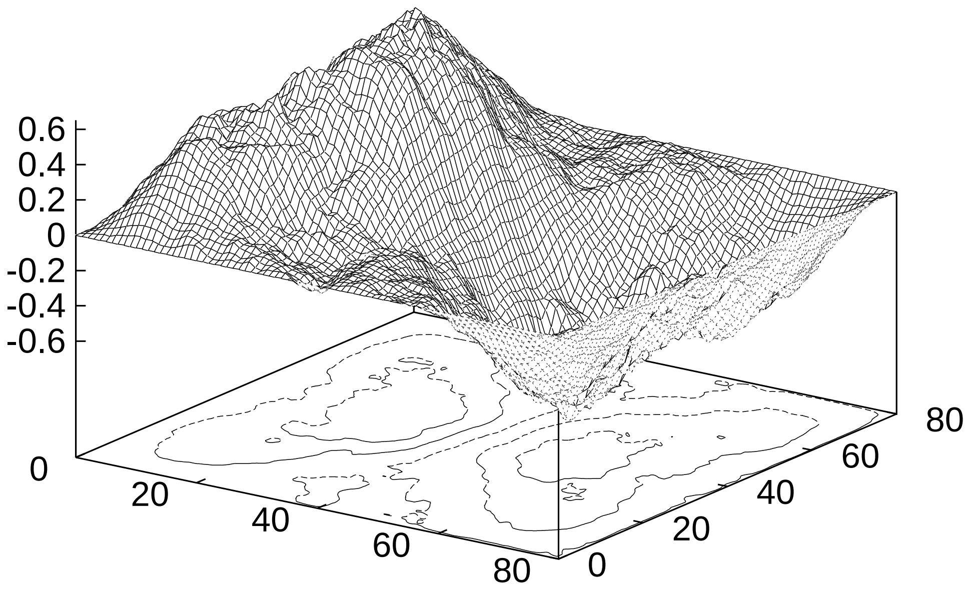

Figure 1.2: cg-iterate after 50 iteration steps, maximum data perturbation 10%, regularization parameter $\alpha = 0.0001$

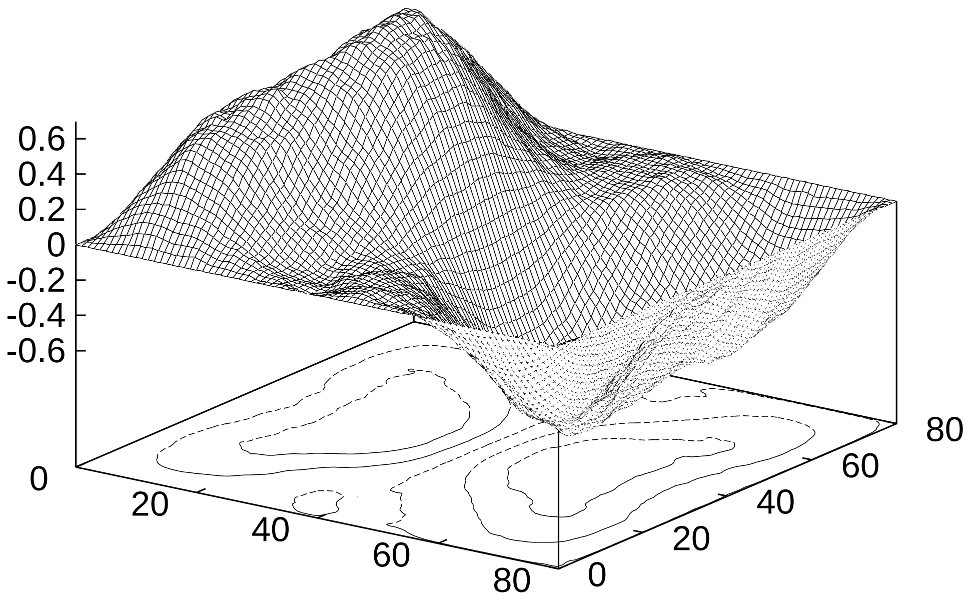

Figure 1.3: multigrid-iterate after 5 iterations, maximum data perturbation 10%, regularization parameter $\alpha = 0.0001$

Nested iteration

Examples above: fairly satisfactory convergence behaviour in the case of perturbed data especially for multigrid but poor convergence properties of all iterative solvers for small regularization parameters α.

Simple practical remedy: nested iteration, i. e., start on the fine grid with a good initial approximation computed on the coarser grid.

Experimental results:

- Full multigrid better than full cg
- Amount of work to obtain a good approximation for the initial function by the icp has nearly the same order of magnitude as the amount of work required for the solution of the direct problem.

Numerical examples (3)

Comparison of full cg and full multigrid for exact data by some pictures

- Initial function for simulation:

$$u_0(x,y) = \left\{\sin(16\pi((x-0.6)^2+(y-0.5)^2)+\tfrac{\pi}{2})+0.2\right\}$$
$$\cdot x^2(1-x)^2 \cdot y^2(1-y)^2 \cdot \mathrm{e}^{-6\sqrt{(x-0.34)^2+(y-0.65)^2}}$$

- icp: $j_{\max} = 5000$ randomly distributed exact data between $t_1 = 0.1$ and $t_2 = 0.7$.
- finest grid: $n = 120$, $\delta t = 0.001$
- regularization parameter $\alpha = 0.000003$ on all grids
- 1 multigrid W-cycle with (1,1,10) (pre/post/coarse) iterations in fmg and a comparable number of 3 cg iterations per grid in fcg, respectively.

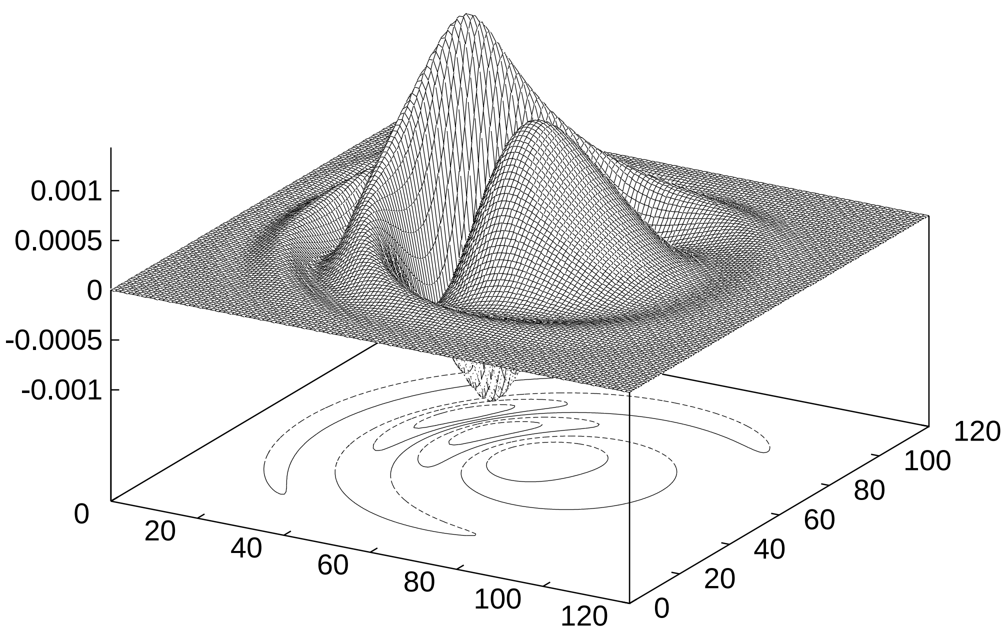

Figure 2.0: Exact inital function on finest grid $n = 120$

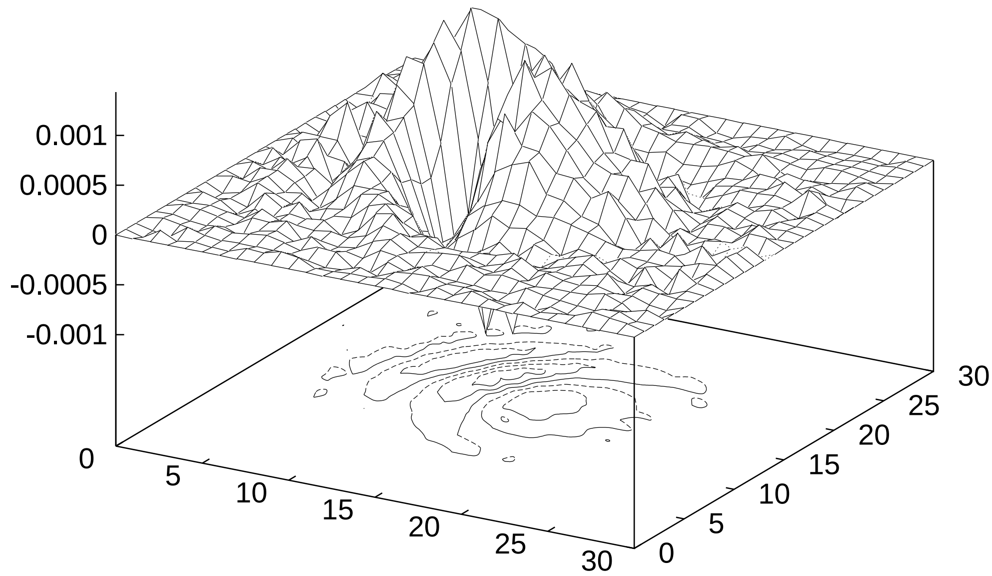

Figure 2.1: Approximation after 5 cg iterations on the coarse start grid with $n = 30$ in the fcg procedure; maximum error to exact initial function 27.1%

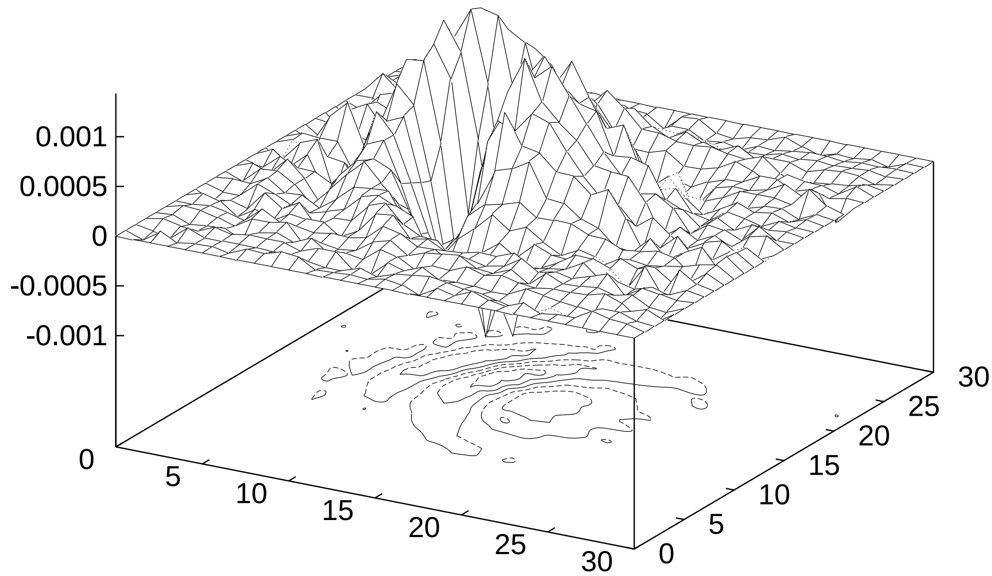

Figure 2.2: Approximation after 5 iterations of the gradient method on the coarse start grid with $n = 30$ in the fmg procedure; maximum error to exact initial function 27.2%

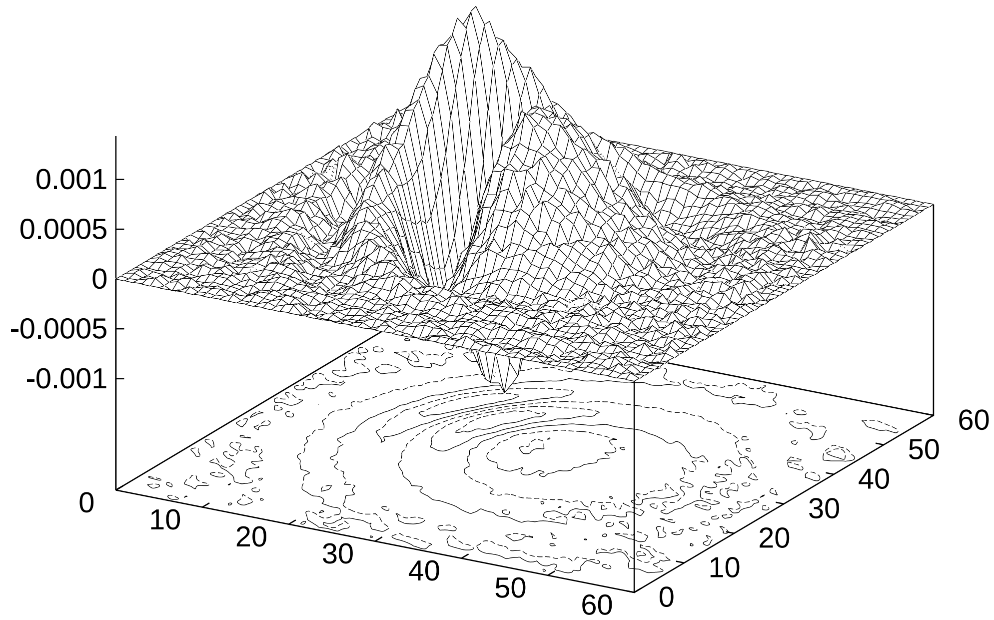

Figure 2.3: Approximation after 3 cg iterations on the second grid with $n = 60$ in the fcg procedure; maximum error to exact initial function 12.5%

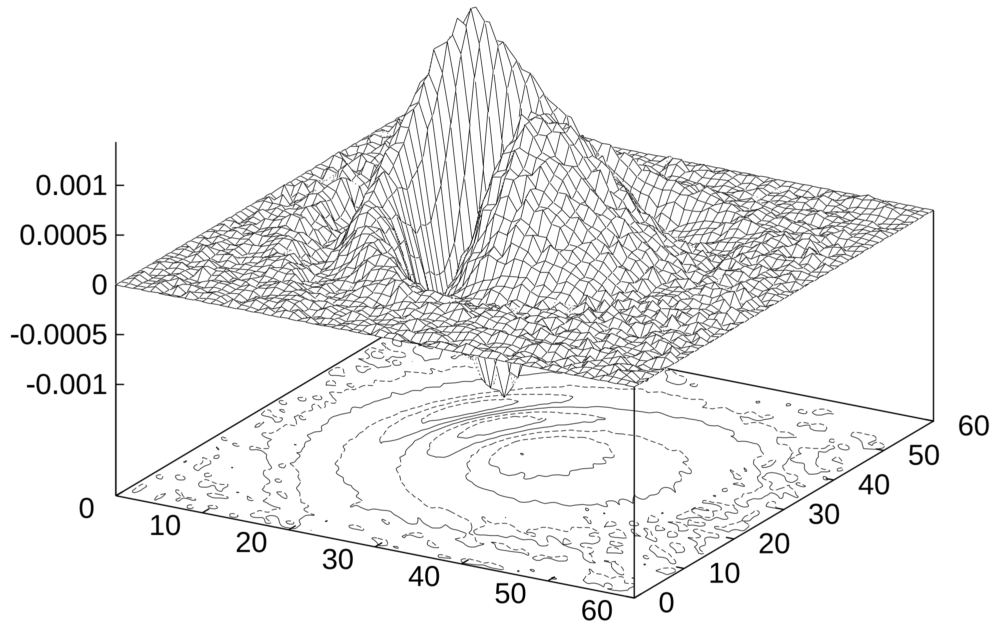

Figure 2.4: Approximation after 1 (1,1,10) W-cycle iteration on the second grid with $n = 60$ in the fmg procedure; maximum error to exact initial function 11.1%

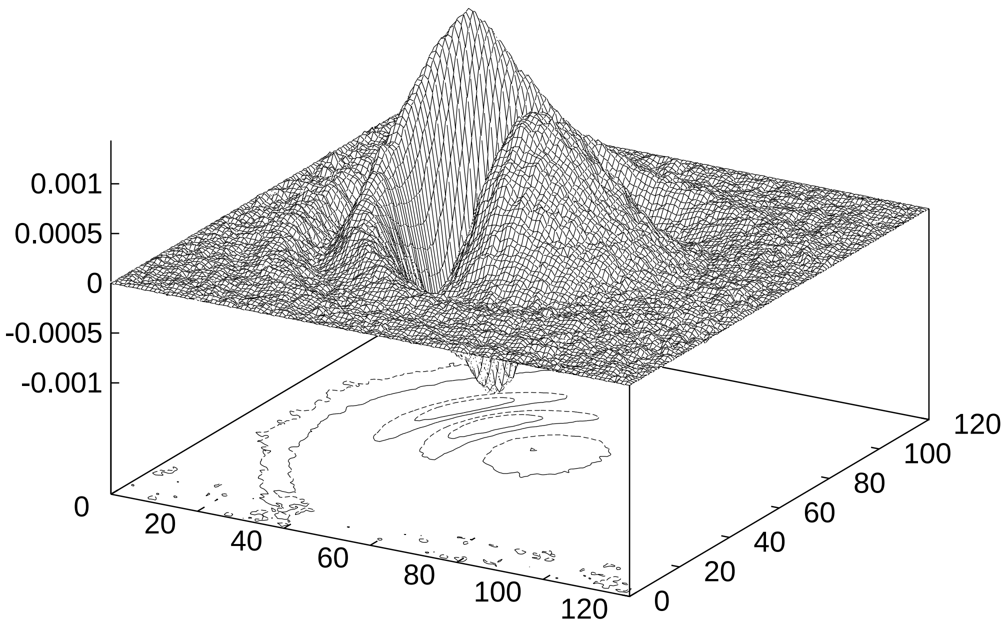

Figure 2.5: Approximation after 6 cg iterations on the finest grid with $n = 120$ in the fcg procedure; maximum error to exact initial function 6.8%

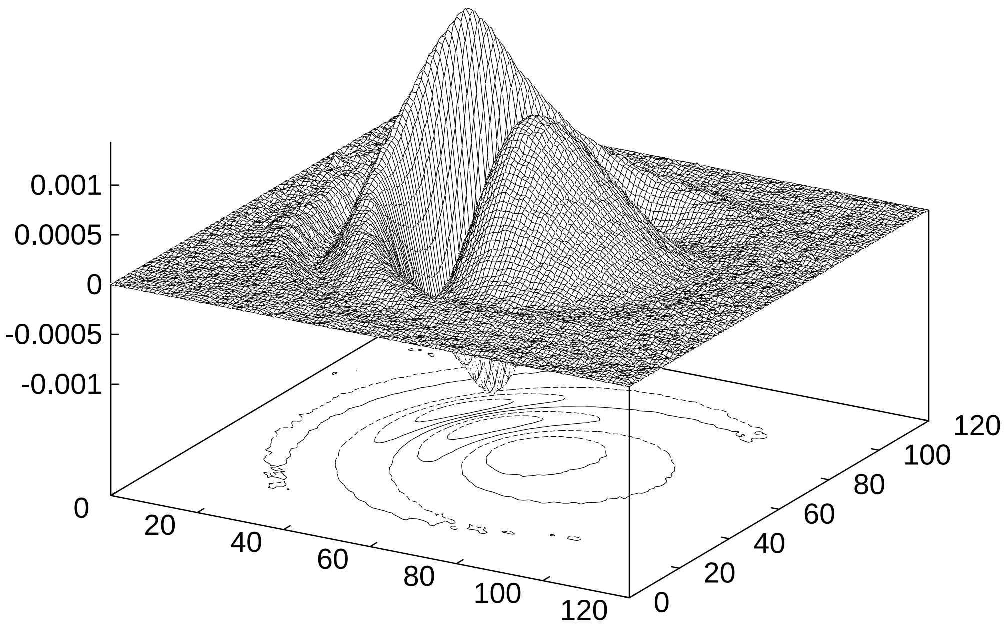

Figure 2.6: Approximation after 2 (1,1,10)-W-cycle iterations on the finest grid with $n = 120$ in the fmg procedure; maximum error to exact initial function 4.0%

Numerical examples (4)

Graphical comparison of full cg and full multigrid for exact data

- Initial function for simulation:

$$u_0(x,y) = \tfrac{1}{2}\sin 5\pi x \sin 2\pi y - \tfrac{1}{5}\sin 8\pi x \sin 4\pi y$$
$$+100x(1-x)(x-0.6)y(1-y)(0.2-y).$$

- icp: $j_{\max} = 6000$ randomly distributed exact data between $t_1 = 0.09$ and $t_2 = 0.34$.

- Notation:

$${}^{n_{\mathrm{f}}}_{\ell}\mathrm{fmg}[i_s, m \cdot (n_{\mathrm{pre}}, n_{\mathrm{post}}, n_{\mathrm{coarse}})X] \quad \text{and} \quad {}^{n_{\mathrm{f}}}_{\ell}\mathrm{fcg}[i_s, j]$$

n_{f}: discretization parameter on the finest grid

ℓ: number of grids in the nested iteration

i_s: number of iterations of the gradient method or cg, respectively, on the starting grid

$m\cdot (n_{\mathrm{pre}}, n_{\mathrm{post}}, n_{\mathrm{coarse}})X$: number of X-cycles ($X = V$ or $X = W$) with n_{pre} iterations in the pre-smoothing step, n_{post} iterations in the post-smoothing step and n_{coarse} iterations on the coarsest grid in the fmg procedure

j: number of cg iterations on the finer grids

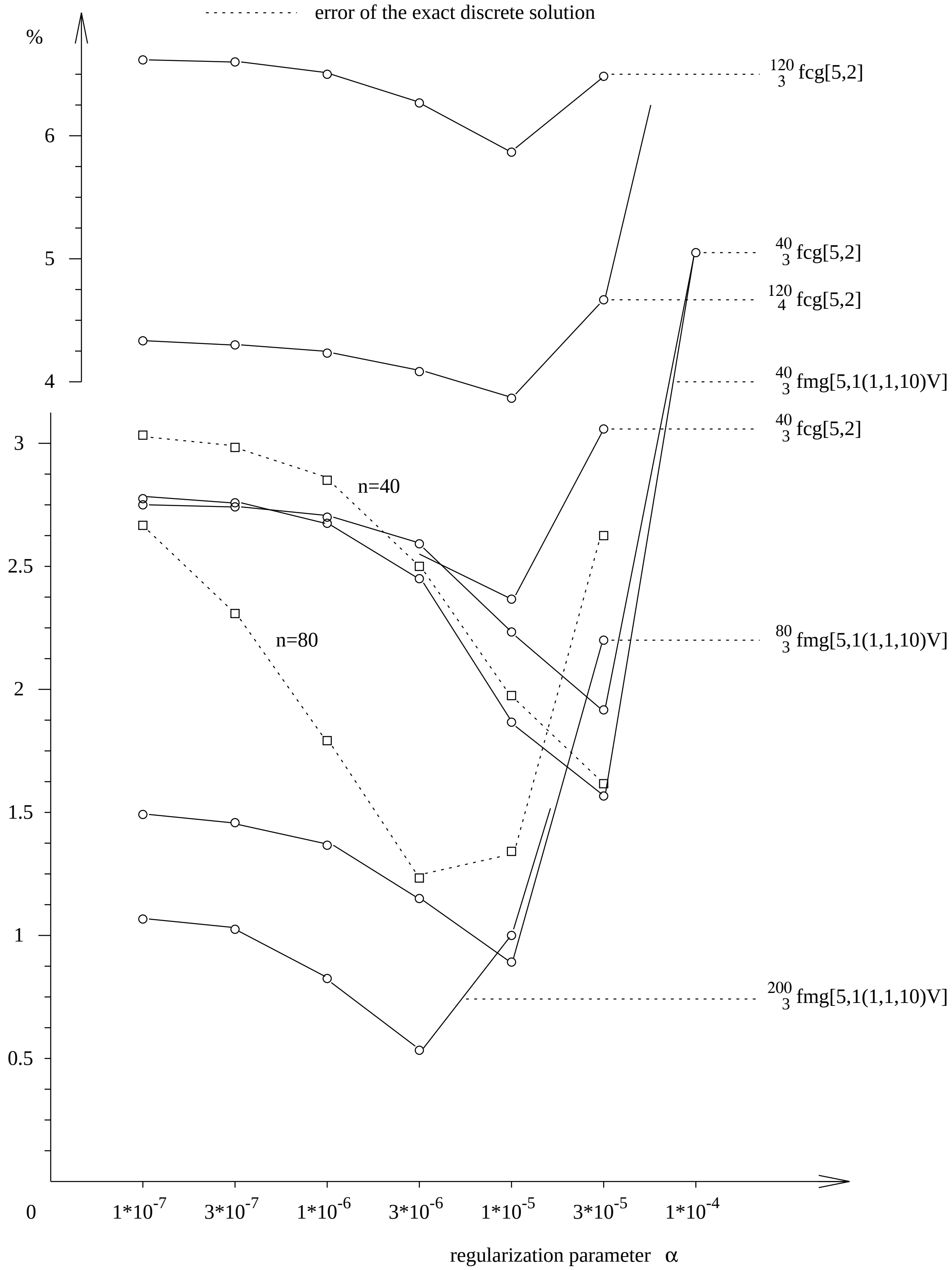

Figure 3: Relative maximum error of the approximations for several regularization and discretization parameters using fmg and fcg

Some outlook

- recently treated: initial control problem for the shallow water equations as a nonlinear hyperbolic system;
  complication at this: computation of the adjoint operator, parameter splitting in the gradient method as smoothing iteration in the full approximation scheme
- further interesting points:
  - mathematically: use of reasonable adjoint operators, modifications of the minimization functional in the regularization term and in the least squares term due to the time-dependent nature of the underlying equation
  - practically: adaptivity in the discretization of the underlying equations, parallelization by grid partitioning will be straight forward (at the moment: parallel multigrid demo based on Fortran 90 and MPI)