On solution of linear problems by the extrapolated Tikhonov method

Uno Hämarik, Reimo Palm, Toomas Raus

September 7, 2009

The problem

- We consider an operator equation

$$Au = f_0,$$

where $A : X \to F$ is a linear bounded operator between real Hilbert spaces, $f_0 \in \mathcal{R}(A) \Rightarrow \exists$ solution $u_* \in H$.
- Instead of exact data f_0, noisy data f are available.
- Knowledge of $\|f_0 - f\|$:
  - Case 1: exact noise level δ: $\|f_0 - f\| \leq \delta$
  - Case 2: approximate noise level δ: $\lim \|f_0 - f\|/\delta \leq C$ as $\delta \to 0$, with unknown constant C
  - Case 3: no information about $\|f_0 - f\|$

Motivation of extrapolation

- Tikhonov approximation: $u_\alpha = (\alpha I + A^*A)^{-1}A^*f$. Extrapolated approximations are linear combinations of u_α with different α. Examples: $v_{2,\alpha} = 2u_\alpha - u_{2\alpha}$, $v_{3,\alpha} = \frac{8}{3}u_{\alpha/2} - 2u_\alpha + \frac{1}{3}u_{2\alpha}$.
- Error estimate: if $\|f - f_0\| \leq \delta$ and

$$u_* \in \mathcal{R}((A^*A)^{p/2}),$$

then for proper α

$$\|u_\alpha - u_*\| \leq \text{const}\ \delta^{p/(p+1)} \qquad (p \leq 2)$$

$$\|v_{2,\alpha} - u_*\| \leq \text{const}\ \delta^{p/(p+1)} \qquad (p \leq 4)$$

$$\|v_{3,\alpha} - u_*\| \leq \text{const}\ \delta^{p/(p+1)} \qquad (p \leq 6)$$

Approximate solutions

- Tikhonov method: $v_{1,\alpha} = u_\alpha = (\alpha I + A^* A)^{-1} A^* f$
- Extrapolated Tikhonov approximations, computed on grid $\alpha_i = q^i$, $i = 0, 1, \ldots$ ($q < 1$, we used $q = 0.9$):

$$v_{2,\alpha_i} = (1-q)^{-1}(u_{\alpha_i} - q u_{\alpha_{i-1}}).$$

- For choice of parameter α_i we also compute

$$v_{3,\alpha_i} = (1-q)^{-2}\left[(1+q)^{-1} u_{\alpha_{i+1}} - q u_{\alpha_i} + q^3 (1+q)^{-1} u_{\alpha_{i-1}}\right],$$
$$v_{4,\alpha_i} = (1-q)^{-3}(1+q)^{-1}\left[(1+q+q^2)^{-1} u_{\alpha_{i+1}} - q u_{\alpha_i} + q^3 u_{\alpha_{i-1}} - q^6 (1+q+q^2)^{-1} u_{\alpha_{i-2}}\right]$$

Rules for choice of parameters in v_{k,α_i} for exact noise level δ

- We compute v_{k,α_i} and $r_{k,i} = Av_{k,\alpha_i} - f$ for $\alpha_0 = 1$, $\alpha_1 = q$, $\alpha_2 = q^2, \ldots$ until some condition is satisfied.
- Approach 1. ME-rule: choose $k = k_{\mathrm{ME}}$ as the first k with

$$d_{\mathrm{ME}}(k) := \frac{(r_{k,0} + r_{k+1,0}, r_{k+1,0})}{2\|r_{k+1,0}\|} \leq \delta.$$

- Approach 2. We choose α_i in v_{k,α_i} ($k \in \{1, 2\}$) by rules:
  - α_{D} is α_i with first i for which $\|r_{k,i}\| \leq \delta$.
  - α_{ME} is α_i with first i for which

$$(r_{k,i}, r_{k+1,i})/\|r_{k+1,i}\| \leq \delta.$$

  - $\alpha_{\mathrm{R2},\tau}$ is α_i with first i for which

$$d_{\mathrm{R2},\tau}(\alpha_i) := \frac{\sqrt{\alpha_i}\,\|v_{k,\alpha_i} - v_{k+1,\alpha_i}\|^2(1 + \alpha\|A\|^{-2})^\tau}{(v_{k,\alpha_i} - v_{k+1,\alpha_i}, v_{k+1,\alpha_i} - v_{k+2,\alpha_i})^{1/2}} \leq b_k\delta$$

with $b_1 = 0.3$, $b_2 = 0.2$; $0 \leq \tau \leq 1$.

Minimum strategy for choice of α in case of exact noise level δ

- $\alpha_{\mathrm{ME}} \geq \alpha_{\mathrm{opt}} := \operatorname{argmin}\{\|u_\alpha - u_*\|, \alpha \geq 0\}$, computations suggested to use $\alpha_{\mathrm{MEe}} = \min(0.5\alpha_{\mathrm{ME}}, 0.6\alpha_{\mathrm{ME}}^{1.07})$, which is good in case $\|f - f_0\| = \delta$
- $\alpha_{\mathrm{R2e}} = 0.5\alpha_{\mathrm{R2},1/2}$ is good in case $\|f - f_0\| < \delta$
- In both cases $\alpha_{\mathrm{MR2e}} = \min(\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}})$ chooses the best of α_{MEe} and α_{R2e}.

Rule for choice of α in case of approximate noise level δ

Rule DM

1) find $\underline{\alpha}$ as the first α_i for which $\sqrt{\alpha_i}\|v_{k,\alpha_i} - v_{k+1,\alpha_i}\| \leq c_1\delta$, $c_1 = \text{const}$ (we used $c_1 = 0.001 \ldots 0.02$);
2) find $\alpha_i = \text{argmin}\, d_{R2,1}(\alpha_i)\alpha_i^{c_2-0.5}$ on $[\underline{\alpha}, 1]$. We used $c_2 = 0.03 \ldots 0.14$. If the first condition is not fulfilled up to $\alpha_i = 10^{-30}$, then $\underline{\alpha} = 10^{-30}$.

Convergence and convergence rate

- Approach 1. ME-rule gives $k = k_{\mathrm{ME}}$ with property $\|v_{k,\alpha_0} - u_*\| < \|v_{k-1,\alpha_0} - u_*\|$ for $k = 1, 2, \ldots, k_{\mathrm{ME}}$, $\|v_{k_{\mathrm{ME}},\alpha_0} - u_*\| \to 0$ $(\delta \to 0)$. If $u_* \in \mathcal{R}((A^*A)^{p/2})$, then $\|v_{k_{\mathrm{ME}},\alpha_0} - u_*\| \leq \mathrm{const}\, \delta^{p/(p+1)}$ for all $p > 0$.
- Approach 2. Convergence $v_{k,\alpha} \to u_*$ $(\delta \to 0)$ is guaranteed for choice of α by rules ME and R2 (and also by rule DM, if $\lim \|f_0 - f\|/\delta \leq C$ as $\delta \to 0$). If $u_* \in \mathcal{R}((A^*A)^{p/2})$, the rules ME and R2 (and DM if $c_1 \geq 0.24$) guarantee $\|v_{k,\alpha} - u_*\| \leq \mathrm{const}\, \delta^{p/(p+1)}$ $(p \leq 2k)$.
- If the parameter choice rule does not use δ, no convergence $v_{k,\alpha} \to u_*$ $(\delta \to 0)$ is guaranteed.

Heuristic rules not using noise level δ

The following rules are modifications of the quasioptimality criterion and the rule from [Brezinski, Rodriguez, Seatzu, 2008, Numer. Algor. 49:85–104].

- **Rules QC, R2C and BRSC**: find α_i as the minimizer of $\psi(\alpha_i)$ (QC: $\psi(\alpha_i) = \|v_{k,\alpha_i} - v_{k+1,\alpha_i}\|$, R2C: $\psi(\alpha_i) = d_{\mathrm{R2,1}}(\alpha_i)$, BRSC: $\psi(\alpha_i) = \|r_{k,i}\|^2/(\alpha_i\|v_{k,\alpha_i}\|)$) on the interval $[\underline{\alpha}, 1]$, where $\underline{\alpha}$ is the largest α_i, for which the value of $\psi(\alpha_i)$ is C times (QC, R2C: $C = 5$; BRSC: $C = 3$) larger than its value at the minimum point.
- **Rules DR21 and BRS1**: choose α_i as the largest local minimizer of functions $\|r_{k,i}\| d_{\mathrm{R2,1}}(\alpha_i)\alpha_i^{0.4}$ and $\|r_{k,i}\|^2/(\alpha\|v_{k,\alpha_i}\|)\alpha_i^{0.7}$, respectively.

Test problems

- Test problems: 10 problems from P. C. Hansen's *Regularization tools* + additional problems `hilbert`, `gauss`, `lotkin`, `moler`, `pascal`, `prolate` from paper [Brezinski, Rodriguez, Seatzu, 2008, Numer. Algor. 49:85–104].
- Besides solution u_* also smoother solution $u_{*,p} = (A^*A)^{p/2}u_*$ with $f_0 = Au_{*,p}$, $p = 2$ was used.
- The problems were normalized, so that norms of the operator and the right hand side were 1.
- For perturbed data we took $f = f_0 + \Delta$, $\|\Delta\| = 0.5, 10^{-1}, \dots, 10^{-6}$ with 10 different perturbations Δ generated by computer.

Data in Tables 1–7

In the following tables we present averages (in Table 1 also maximums) of error ratios $\|u_\alpha - u_*\|/e_{\text{opt}}$, where $e_{\text{opt}} = \min\{\|u_\alpha - u_*\| : \alpha \geq 0\}$.

- In Tables 1–5 we use $\delta = d\|f_0 - f\|$ with $d = 1, 1.3, 2$ (in Table 1) or $d = 0.01, 0.1, 0.5, 1, 2, 10, 100$ (in Tables 2, 3; in Table 4 additionally $d = 0.3, 30$ and in Table 5 $d = 0.03, 0.3, 4, 30$); $d \geq 1$ corresponds to overestimation of noise level. Rules of Tables 6, 7 do not use δ.
- Columns vR2e, vMEe, vMR2e, vQC, vR2C, vBRSC, vQ1, vBRS1 show averages and maximums of error ratios $\|v_{2,\alpha} - u_*\|/e_{\text{opt}}$ for rules R2e, MEe, MR2e, QC, R2C, BRSC, Q1, BRS1.

Table 1, averages and maximums

	d	p	D	MEe	R2e	MR2e	vMEe	vR2e	vMR2e
Averages	1	0	1.24	1.18	1.38	1.18	1.17	1.45	1.29
	1	2	2.68	1.20	1.17	1.13	0.69	0.79	0.77
	1.3	0	1.78	1.66	1.44	1.43	1.60	1.41	1.39
	1.3	2	3.66	3.26	1.25	1.25	1.41	0.77	0.77
	2	0	2.15	1.96	1.56	1.56	1.87	1.50	1.49
	2	2	4.72	4.43	1.53	1.53	2.06	0.78	0.78
Maximums	1	0	5.82	5.18	16	4.96	5.39	25	25
	1	2	25	119	46	29	399	595	595
	1.3	0	20	19	16	16	21	19	19
	1.3	2	4e3	3e3	241	241	723	500	500
	2	0	25	22	17	17	21	19	19
	2	2	5e3	3e3	844	844	850	277	277

Table 2. Rule DM, $c_1 = 0.002$, $c_2 = 0.03$, $p = 0$, $k = 1, 2$

Problem\\d	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
baart	1.46	1.46	1.46	1.46	1.49	1.69	2.51	1.53	1.53	1.53	1.54	1.56	1.71	2.55
deriv2	1.56	1.56	1.34	1.08	1.08	1.07	1.25	7.93	5.04	1.57	1.57	1.56	1.09	1.20
foxgood	2.02	2.02	2.02	2.02	2.02	1.84	5.88	2.55	2.55	2.55	2.55	2.55	2.27	6.26
gravity	1.12	1.12	1.12	1.12	1.12	1.11	1.62	1.11	1.11	1.11	1.11	1.11	1.08	1.62
heat	1.66	1.16	1.16	1.10	1.10	1.10	1.17	6.42	2.37	1.15	1.15	1.10	1.10	1.17
i_laplace	1.16	1.16	1.16	1.16	1.16	1.16	1.44	1.19	1.19	1.19	1.19	1.19	1.19	1.44
phillips	1.11	1.11	1.11	1.11	1.11	1.11	1.36	1.08	1.08	1.08	1.08	1.08	1.08	1.34
shaw	1.39	1.39	1.39	1.39	1.39	1.46	2.06	1.45	1.45	1.45	1.45	1.46	1.49	2.12
spikes	1.03	1.03	1.03	1.03	1.03	1.03	1.05	1.04	1.04	1.04	1.04	1.04	1.04	1.06
wing	1.42	1.42	1.42	1.42	1.42	1.47	1.54	1.42	1.42	1.42	1.42	1.43	1.48	1.54
gauss	1.16	1.16	1.16	1.16	1.16	1.16	1.56	1.18	1.18	1.18	1.18	1.18	1.16	1.58
hilbert	1.43	1.43	1.43	1.43	1.43	1.47	2.13	1.56	1.56	1.56	1.57	1.56	1.64	2.32
lotkin	2.41	2.41	2.41	2.41	2.41	2.43	3.80	1.79	1.79	1.79	1.79	1.79	1.77	2.63
moler	3.28	1.84	1.66	1.56	1.45	1.35	1.71	18.34	4.73	2.92	2.08	1.84	1.56	1.61
pascal	1.05	1.05	1.05	1.05	1.06	1.06	1.06	1.05	1.05	1.05	1.06	1.06	1.06	1.06
prolate	1.36	1.36	1.36	1.36	1.33	1.35	2.21	1.46	1.46	1.46	1.46	1.43	1.57	2.34
Average	1.54	1.42	1.39	1.37	1.36	1.37	2.02	3.19	1.91	1.50	1.45	1.43	1.39	1.99

Table 3. Rule DM, $c_1 = 0.002$, $c_2 = 0.03$, $p = 2$, $k = 1, 2$

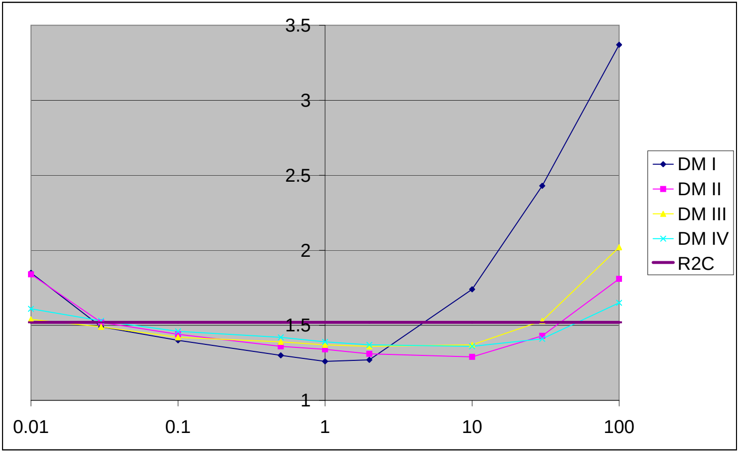
Problem\\d	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
baart	1.93	1.93	1.93	1.84	1.74	1.33	3.33	1.24	1.24	1.24	1.17	1.08	0.87	2.13
deriv2	1.22	1.22	1.22	1.22	1.22	1.22	1.43	15.85	0.68	0.68	0.68	0.68	0.67	0.94
foxgood	1.55	1.55	1.55	1.55	1.55	1.26	3.23	0.88	0.88	0.88	0.88	0.89	0.60	1.77
gravity	1.33	1.33	1.33	1.33	1.33	1.25	1.75	0.70	0.70	0.70	0.70	0.70	0.65	0.87
heat	1.21	1.21	1.21	1.21	1.21	1.21	1.15	0.79	0.79	0.79	0.79	0.79	0.79	0.82
i_laplace	1.33	1.33	1.33	1.33	1.33	1.29	1.82	0.77	0.77	0.77	0.77	0.77	0.74	1.05
phillips	1.21	1.21	1.21	1.21	1.21	1.21	1.37	0.58	0.58	0.58	0.58	0.58	0.58	0.72
shaw	1.49	1.49	1.49	1.49	1.49	1.26	2.02	0.92	0.92	0.92	0.92	0.86	0.76	1.20
spikes	1.42	1.42	1.42	1.42	1.42	1.26	2.38	0.89	0.89	0.89	0.89	0.89	0.79	1.58
wing	2.20	2.20	2.20	1.85	1.62	1.28	3.85	1.34	1.34	1.34	1.13	1.05	0.77	2.59
gauss	1.30	1.30	1.30	1.30	1.30	1.25	1.69	0.70	0.70	0.70	0.70	0.70	0.68	0.86
hilbert	1.50	1.50	1.50	1.50	1.50	1.26	2.47	1.04	1.04	1.04	1.04	1.04	0.87	1.72
lotkin	1.46	1.46	1.46	1.46	1.46	1.33	2.81	0.95	0.95	0.95	0.95	0.95	0.80	1.71
moler	1.23	1.23	1.23	1.23	1.23	1.21	1.92	7.61	0.62	0.62	0.62	0.62	0.60	1.23
pascal	6.04	6.04	5.22	4.38	3.60	3.05	18.33	3.71	3.69	3.54	3.42	3.25	2.63	11.13
prolate	1.32	1.32	1.32	1.32	1.32	1.16	1.81	0.71	0.71	0.71	0.71	0.71	0.65	0.78
Average	1.73	1.73	1.68	1.60	1.53	1.36	3.21	2.42	1.03	1.02	1.00	0.97	0.84	1.94

Table 4. Averages of error ratios over all problems for rule DM for different c_1, c_2; $p = 0$

				$d = \delta/\|f_0 - f\|$								
k	Nr	c_1	c_2	0.01	0.03	0.1	0.5	1	2	10	30	100
1	I	0.02	0.14	1.85	1.49	1.40	1.30	1.26	1.27	1.74	2.43	3.37
1	II	0.002	0.07	1.84	1.52	1.44	1.36	1.34	1.31	1.29	1.43	1.81
1	III	0.002	0.03	1.54	1.49	1.42	1.39	1.37	1.36	1.37	1.53	2.02
1	IV	0.001	0.03	1.61	1.53	1.46	1.42	1.39	1.37	1.36	1.41	1.65
2	I	0.02	0.14	3.73	2.41	1.68	1.38	1.35	1.33	1.81	2.37	3.33
2	II	0.002	0.07	5.69	3.15	2.20	1.55	1.46	1.40	1.35	1.48	1.87
2	III	0.002	0.03	3.19	2.52	1.91	1.50	1.45	1.43	1.39	1.54	1.99
2	IV	0.001	0.03	5.43	2.72	2.17	1.57	1.50	1.45	1.40	1.44	1.68

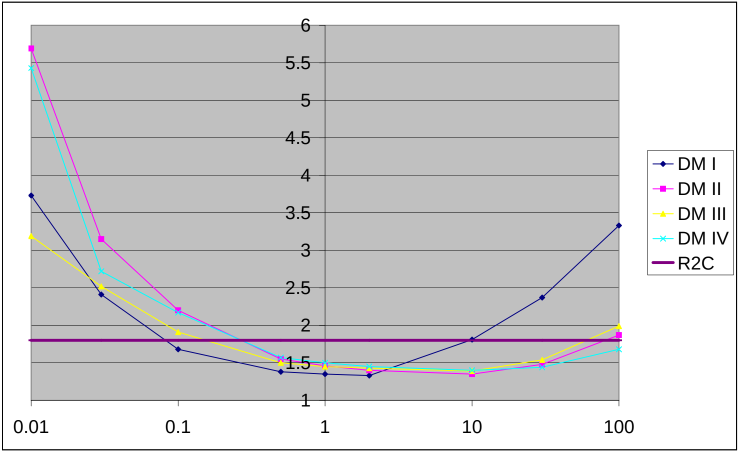
In case of fewer information about noise level lower lines for parameters c_1, c_2 are recommended. In Table 6 a δ-free rule R2C gives average 1.52 for $k = 1$ and 1.80 for $k = 2$. Hence DM is superior over R2C for $k = 1$, $d \in [0.03, 30]$ and for $k = 2$, $d \in [0.1, 10]$.

Error ratios for rules R2C and DM: $p = 0$, $k = 1$



$$d = \delta / \|f_0 - f\|$$

Error ratios for rules R2C and DM: $p = 0$, $k = 2$



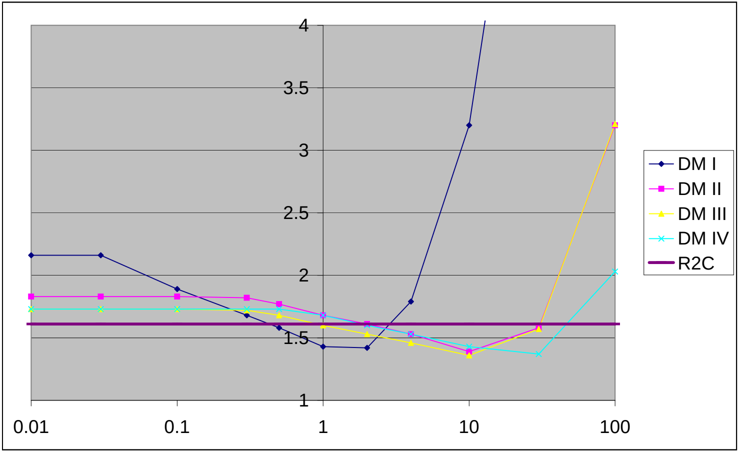
$$d = \delta / \|f_0 - f\|$$

Table 5. Averages of error ratios over all problems for rule DM for different c_1, c_2; $p = 2$

				$d = \delta/\|f_0 - f\|$										
k	Nr	c_1	c_2	0.01	0.03	0.1	0.3	0.5	1	2	4	10	30	100
1	I	0.02	0.14	2.16	2.16	1.89	1.68	1.58	1.43	1.42	1.79	3.20	6.85	12.91
1	II	0.002	0.07	1.83	1.83	1.83	1.82	1.77	1.68	1.61	1.53	1.39	1.58	3.20
1	III	0.002	0.03	1.73	1.73	1.73	1.72	1.68	1.60	1.53	1.46	1.36	1.57	3.21
1	IV	0.001	0.03	1.73	1.73	1.73	1.73	1.73	1.68	1.60	1.53	1.43	1.37	2.03
2	I	0.02	0.14	2.92	1.29	1.20	1.08	0.99	0.89	0.79	0.94	1.94	3.91	7.65
2	II	0.002	0.07	6.36	2.29	1.12	1.11	1.09	1.06	1.01	0.96	0.85	0.81	1.94
2	III	0.002	0.03	2.42	1.03	1.03	1.02	1.02	1.00	0.97	0.93	0.84	0.81	1.94
2	IV	0.001	0.03	4.98	1.85	1.03	1.03	1.03	1.02	1.00	0.97	0.90	0.80	1.13

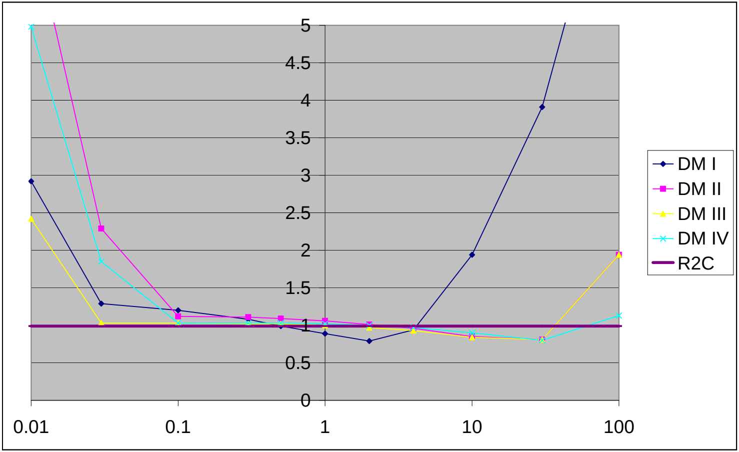
In case of fewer information about noise level lower lines for parameters c_1, c_2 are recommended. In Table 7 a δ-free rule R2C gives average 1.61 for $k = 1$ and 0.99 for $k = 2$. Hence DM is superior over R2C for $k = 1$, if $d \in [0.5, 2]$ or $d \in [1, 30]$ and for $k = 2$, if $d \in [0.5, 4]$ or $d \in [2, 30]$.

Error ratios for rules R2C and DM: $p = 2$, $k = 1$



$$d = \delta / \|f_0 - f\|$$

Error ratios for rules R2C and DM: $p = 2$, $k = 2$



$$d = \delta / \| f_0 - f \|$$

Table 6. Rules not using noise level, $p = 0$, $k = 1, 2$

Problem	QC	R2C	BRSC	DR21	BRS1	vQC	vR2C	vBRSC	vDR21	vBRS1
baart	1.55	1.52	2.60	1.81	1.91	1.90	1.89	2.63	1.70	2.36
deriv2	1.61	1.57	1.35	1.55	1.62	1.62	1.59	1.36	1.55	1.64
foxgood	2.16	2.11	5.24	1.71	3.20	2.43	2.38	7.09	2.39	3.88
gravity	1.12	1.10	2.08	1.16	1.32	1.11	1.09	2.23	1.72	1.55
heat	1.33	1.28	1.35	1.88	1.21	1.29	1.27	1.42	2.10	1.17
i_laplace	1.19	1.17	1.87	1.13	1.38	1.19	1.18	1.85	1.27	1.42
phillips	1.07	1.08	1.61	1.26	1.12	1.09	1.08	1.78	1.82	1.23
shaw	1.43	1.45	2.25	1.43	1.59	1.53	1.52	2.44	1.41	1.73
spikes	1.04	1.04	1.06	1.04	1.03	1.05	1.05	1.07	1.04	1.05
wing	1.43	1.42	1.55	1.47	1.47	1.42	1.42	1.88	1.69	1.84
gauss	1.18	1.16	1.87	1.19	1.30	1.18	1.17	2.01	1.52	1.49
hilbert	1.74	1.89	2.63	1.27	1.64	1.96	1.93	3.14	1.27	2.14
lotkin	3.29	3.26	2.92	2.99	1.83	3.27	3.26	3.89	3.26	2.07
moler	1.87	1.84	1.96	1.82	1.88	5.32	5.29	5.48	15.35	5.21
pascal	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
prolate	1.35	1.34	1.56	1.29	1.29	1.67	1.66	2.27	1.45	1.81
Average	1.53	1.52	2.06	1.50	1.55	1.82	1.80	2.60	2.54	1.98

Table 7. Rules not using noise level, $p = 2$, $k = 1, 2$

Problem	QC	R2C	BRSC	DR21	BRS1	vQC	vR2C	vBRSC	vDR21	vBRS1
baart	1.74	1.92	2.69	1.92	1.91	1.10	1.16	2.44	2.62	1.27
deriv2	1.06	1.19	3.10	6.32	19.94	0.61	0.66	1.07	8.14	0.66
foxgood	1.28	1.40	3.44	2.15	2.56	0.82	0.86	1.90	3.14	0.90
gravity	1.13	1.27	2.24	2.60	2.45	0.64	0.69	0.90	3.72	0.56
heat	1.06	1.19	2.41	3.21	3.45	0.72	0.78	0.97	4.83	0.69
i_laplace	1.14	1.30	1.99	2.27	1.91	0.70	0.75	1.12	3.17	0.67
phillips	1.07	1.20	2.55	3.01	3.46	0.54	0.57	0.75	4.58	0.52
shaw	1.28	1.42	2.17	2.35	1.97	0.81	0.86	1.28	3.21	0.73
spikes	1.20	1.40	2.55	2.18	1.96	0.82	0.87	1.99	3.02	1.02
wing	1.77	1.88	3.29	2.01	1.84	1.19	1.23	3.07	2.55	1.55
gauss	1.11	1.25	2.03	2.65	2.28	0.63	0.67	0.85	3.65	0.57
hilbert	1.22	1.39	2.23	2.14	1.89	0.85	0.91	1.82	2.96	1.03
lotkin	1.30	1.41	2.64	2.10	2.08	0.91	0.92	1.98	2.87	1.19
moler	1.06	1.21	3.41	2.92	6.71	0.57	0.62	1.44	4.67	0.76
pascal	4.97	5.14	27.94	5.10	8.09	3.62	3.64	14.64	5.94	4.41
prolate	1.15	1.23	1.63	2.34	2.17	0.66	0.68	0.62	3.10	0.60
Average	1.47	1.61	4.14	2.83	4.04	0.95	0.99	2.30	3.89	1.07