# On choice of the regularization parameter in ill-posed problems with rough estimate of the noise level of the data

U. Hämarik, R. Palm, T. Raus

University of Tartu, Estonia

Linköping, April 4, 2013

# Contents

# Problem and Tikhonov method

- We consider linear ill-posed problems

$$Ax = y_*, \qquad y_* \in \mathcal{R}(A),$$

where $A\colon X \to Y$ is a linear continuous operator between Hilbert spaces. The range $\mathcal{R}(A)$ may be non-closed and the kernel $\mathcal{N}(A)$ may be non-trivial.

- Assume that instead of exact data $y_*$ only its approximation $y$ is available.
- For approximation of the minimum norm solution $x_*$ of the problem $Ax = y_*$ we use the Tikhonov regularization method

$$x_\alpha = (\alpha I + A^*A)^{-1}A^*y.$$

# Other methods

- Iterated Tikhonov method: take $x_0 = x_{0;\alpha} \in X$ and compute $x_\alpha := x_{m;\alpha}$ iteratively from the equations

$$\alpha x_{n;\alpha} + A^* A x_{n;\alpha} = \alpha x_{n-1;\alpha} + A^* y \qquad (n = 1, ..., m). \qquad (1)$$

- Extrapolated Tikhonov method: take Tikhonov approximations $x_{\alpha_1}$, $\ldots$, $x_{\alpha_m}$ with different parameters $\alpha_1, \ldots, \alpha_m$ and compute

$$x_{\alpha_1,\ldots,\alpha_m} = \sum_{i=1}^{m} d_i x_{\alpha_i}, \qquad d_i = \prod_{j=1, j\neq i}^{m} (1 - \alpha_i/\alpha_j)^{-1}. \qquad (2)$$

  We use $x_\alpha := x_{\alpha_1,\ldots,\alpha_m}$ with $\alpha_n = \alpha r^{(m+1)/2-n}$, $n = 1, ..., m$; $r > 1$.
- If $A = A^* \geq 0$ then Lavrentiev method $x_\alpha = (\alpha I + A)^{-1} y$ may also be used.
- Landweber iteration method

$$x_n = x_{n-1} - \mu A^*(Ax_n - y), \quad \mu \in (0, 1/\|A^* A\|), \quad n = 1, 2, \ldots$$

# Choice of the regularization parameter is a compromize between accuracy and stability

Regularization parameter: $\alpha > 0$ in Tikhonov method and in its iterated and extrapolated versions (for these 3 methods common name T-method is used)
$n \in \mathbb{N}$ in Landweber method

Conflict of interests: Approximation vs Stability

| $\alpha$ small, $n$ large | $\alpha$ large, $n$ small |
|---|---|
| good approximation<br>bad stability | bad approximation<br>good stability |

# Information about noise level

We consider three cases of knowledge about noise level for $\|y - y_*\|$:

- Case 1: exact noise level $\delta$: $\|y - y_*\| \leq \delta$.
- Case 2: no information about $\|y - y_*\|$.
- Case 3: approximate noise level: $\delta$ is given but it is not known whether the inequality $\|y - y_*\| \leq \delta$ holds or not. For example, it may be known that with high probability $\delta / \|y - y_*\| \in [1/10, 10]$. This very useful information should be used for choice of $\alpha = \alpha(\delta)$ in T-method and $n = n(\delta)$ in Landweber method.

# General remarks on rules for choice of the regularization parameters $\alpha$ and $n$

- Rules for the Case 1 (discrepancy principle, etc.) need exact noise level: rules fail for very small underestimation of the noise level and give much large error $\|x_\alpha - x_*\|$ and $\|x_n - x_*\|$ than for optimal parameters already for 10% overestimation.
- Heuristic rules for the Case 2 (L-curve, GCV etc) do not guarantee the convergence $x_\alpha \to x_*$ and $x_n \to x_*$ for $\|y - y_*\| \to 0$.
- Our rules for the Case 3 guarantee $x_\alpha \to x_*$ and $x_n \to x_*$ as $\delta \to 0$, if $\lim_{\delta \to 0} \frac{\|y - y_*\|}{\delta} \leq \text{const}$.

In the following we consider rules for the choice of the regularization parameters if an estimate $\delta$ for the noise level $\|y - y_*\|$ and an estimate for $\rho := \delta / \|y - y_*\|$ about the accuracy of the estimate $\delta$ are given.

# Rules for exact or very slightly overestimated noise level

Case 1) $\|y - y_*\| \le \delta$ where $\rho := \delta/\|y - y_*\|$ is 1 or only slightly larger.
**Discrepancy principle (D)** chooses a constant $C \ge 1$ and in the T-method the parameter $\alpha = \alpha_{\mathrm{D}}$ as the solution of the equation $\|Ax_\alpha - y\| = C\delta$, in the Landweber method the stopping index $n_{\mathrm{D}}$ as the first $n$ with $\|r_n\| \le C\delta$, $r_n := Ax_n - y$.
**Monotone error rule (ME-rule)** chooses in the T-method the parameter $\alpha = \alpha_{\mathrm{ME}}$ as the solution of the equation $\|B_\alpha(Ax_\alpha - y)\|^2/\|B_\alpha^2(Ax_\alpha - y)\| = \delta, \quad B_\alpha := \sqrt{\alpha}(\alpha I + AA^*)^{-1/2}$, and in the Landweber method the stopping index $n_{\mathrm{ME}}$ as the first $n$ with $(r_n + r_{n+1}, r_n)/(2\|r_n\|) \le \delta$.
The name ME-rule is justified by the property $\frac{d}{d\alpha}\|x_\alpha - x_*\|^2 \ge 0$ for each $\alpha \in [\alpha_{\mathrm{ME}}, \infty)$ in the T-method and property $\|x_n - x_*\| \le \|x_{n-1} - x_*\|$ for all $n = 1,\ 2,\ \ldots,\ n_{\mathrm{ME}}$ in the Landweber method. Extensive numerical experiments suggest to use the post-estimated parameters of the **MEe-rule** $\alpha_{\mathrm{MEe}} = \alpha_{\mathrm{ME}}/2.3$ and $n_{\mathrm{MEe}} = \mathrm{round}(2.3 n_{\mathrm{ME}})$, instead the parameters $\alpha_{\mathrm{ME}}$ and $n_{\mathrm{ME}}$, respectively. In average of extensive numerical experiments $\|x_{\alpha_{\mathrm{ME}}} - x_*\| \approx 1.2\|x_{\alpha_{\mathrm{MEe}}} - x_*\|$

# Family of rules for T-method for approximate noise level

The estimate $\delta$ of the noise level $\|y - y_*\|$ is given, e.g. $\rho := \delta/\|y - y_*\| \in [0.3, 10]$. We propose family of rules. Fix $q$, $l$, $k$ such that $2q, 2k, 2l \in N$, $l \geq 0$, $k \geq l/q$ and $\underline{q} \leq q < \infty$, where $\underline{q} = (2m+1)/(m+1)$ for the T-method ($\underline{q} = 3/2$ for the Tikhonov method) and $\underline{q} = 2$ for the Landweber method.

**R-rule for T-method** Choose $\alpha = \alpha(\delta)$ as the largest solution of

$$d(\alpha \mid q, l, k) := \frac{\kappa(\alpha)\|D_\alpha^k B_\alpha (Ax_\alpha - y)\|^{q/(q-1)}}{\|D_\alpha^l B_\alpha^{2q-2}(Ax_\alpha - y)\|^{1/(q-1)}} = b\delta,$$

where $B_\alpha = \sqrt{\alpha}(\alpha I + AA^*)^{-1/2}$, $D_\alpha = \alpha^{-1} AA^* B_\alpha^2$,

$$b \approx \left(\frac{3}{2}\right)^{\frac{3}{2}} \frac{k^k}{(k+3/2)^{k+3/2}} \left(\frac{k^k (l+3/2)^{l+3/2}}{l^l (k+3/2)^{k+3/2}}\right)^{\frac{1}{q-1}},$$

$\kappa(\alpha) = 1$, if $k = l/q$, and $\kappa(\alpha) = (1 + \alpha\|A\|^{-2})^{\frac{kq-l+q/2}{q-1}}$, if $k > l/q$. Note that if $k > l/q$, then $\kappa(\alpha) \to 1$, as $\alpha \to 0$.

Denote this rule by R($q$, $l$, $k$).

# Examples of this family of rules

- Modified discrepancy principle (Raus 1985, Gfrerer 1987): $q = 3/2$, $l = k = 0$
- Monotone error rule (Tautenhahn, Hämarik 1999): $q = 2$, $l = k = 0$
- Rule R1 (Raus 1992): $q = 3/2$, $k = l > 0$
- Balancing principle (Mathé, Pereverzev 2003) can be considered as an approximate variant of rule R1 with $k = 1/2$.

# Family of rules for the Lavrentiev method

Fix the parameters $q$, $k$: $4/3 \le q < \infty$, $k \ge 0$, $2k \in N$, $3q \in N$. Let the constant $b > 1$ if $k = 0$ and $b > 0$ if $k > 0$. Choose the regularization parameter $\alpha = \alpha(\delta)$ as the largest solution of the equation

$$d(\alpha \mid q, k) := \frac{\kappa_\alpha \| D_\alpha^k B_\alpha (A x_\alpha - y) \|^{q/(q-1)}}{\| B_\alpha^{3q/2-1} (A x_\alpha - y) \|^{1/(q-1)}} = b\delta,$$

where $B_\alpha = \alpha(\alpha I + A)^{-1}$, $D_\alpha := A(\alpha I + A)^{-1}$ and

$$\kappa_\alpha = (1 + \alpha \|A\|^{-1})^{\frac{kq + s_0 q/2}{q-1}}, \qquad s_0 = \begin{cases} 0, & \text{if } k = 0, \\ 1, & \text{if } k > 0. \end{cases}$$

# Family of rules for the Landweber method

Choose the stopping index $n = n_{\mathrm{R}}$ as the first index with

$$d(n \mid q, l, k) := \kappa(n) \frac{\|D_n^k(Ax_n - y)\|^{q/(q-1)}}{\|D_n^l(Ax_n - y)\|^{1/(q-1)}} \leq b\delta.$$

Here $D_n := nAA^*$ and
$\kappa(n) = 1$, if $k = l/q$, and
$\kappa(n) = (1 + n^{-1}\|A\|^{-2})^{(kq-l+q/2)/(q-1)}$, if $k > l/q$.

# Existence of solution for family of rules

1. If $k > l/q$ ($q \geq \underline{q}$, $l \geq 0$), then for every $b = \text{const} > 0$ there exist a solution of the equation $d(\alpha \mid q, l, k) = b\delta$ in T-method and a stopping index satisfying $d(n \mid q, l, k) \leq b\delta$ in Landweber method, because $\lim_{\alpha\to\infty} d(\alpha \mid q, l, k) = \infty$ and $\lim_{\alpha\to 0} d(\alpha \mid q, l, k) = 0$, $\lim_{n\to\infty} d(n \mid q, l, k) = 0$.
2. If $k = l/q$ ($q \geq \underline{q}$, $l \geq 0$), $b \geq b_0(q, l, k)$ and $\|y - y_*\| \leq \delta$, then in the T-method the solution of the equation $d(\alpha \mid q, l, k) = b\delta$ exists and in Landweber method there exists a stopping index satisfying $d(n \mid q, l, k) \leq b\delta$.

Results of this and the next slide hold, if in formulations the T-method is replaced by the Lavrentiev method and $l = 0$.

# Convergence and stability

- **Convergence.** Let $\underline{q} \leq q < \infty$, $l \geq 0$, $k \geq l/q$. Let in T-method the parameter $\alpha = \alpha(\delta)$ be the solution of the equation $d(\alpha \mid q, l, k) = b\delta$, $b > b_0(q, l, k)$ or in Landweber method parameter $n = n(\delta)$ stopping index from the condition $d(n \mid q, l, k) \leq b\delta$. If $\|y - y_*\| \leq \delta$, then $\|x_\alpha - x_*\| \to 0$ $(\delta \to 0)$ and $\|x_n - x_*\| \to 0$ $(n \to \infty)$.
- **Stability** (with respect to the inaccuracy of the noise level). Let $\underline{q} \leq q < \infty$, $l \geq 0$, $k > l/q$. Let in the T-method the parameter $\alpha(\delta)$ be the **largest** solution of the equation $d(\alpha \mid q, l, k) = b\delta$ and in Landweber method $n(\delta)$ be the first index satisfying the condition $d(n \mid q, l, k) \leq b\delta$. If $\|y - y_*\| / \delta \leq c =$ const in the process $\delta \to 0$, then $\|x_\alpha - x_*\| \to 0$ and $\|x_n - x_*\| \to 0$.

Under information $\rho \in [1, 5]$ on the accuracy of the noise level (no under-estimation) we recommend the **Me-rule** in the T-method: choose $\alpha_{\mathrm{Me}} = \min(\alpha_{\mathrm{MEe}}, 1.4\alpha_R)$, where $\alpha_{\mathrm{R}}$ is parameter from rule $R(3/2, 1/2, 2)$ with $b = 0.25$.

# Quasioptimality in T-method

Let $\underline{q} \leq q < \infty$, $0 \leq l \leq q/2$, $l/q \leq k \leq l$. Let the parameter $\alpha(\delta)$ be the **smallest** solution of the equation $d(\alpha \mid q, l, k) = b\delta$. Then the rule is **quasioptimal**:

$$\|x_\alpha - x_*\| \leq C(b) \inf_{\alpha \geq 0} \left\{ \|x_\alpha^+ - x_*\| + \gamma_* \frac{\delta}{\sqrt{\alpha}} \right\},$$

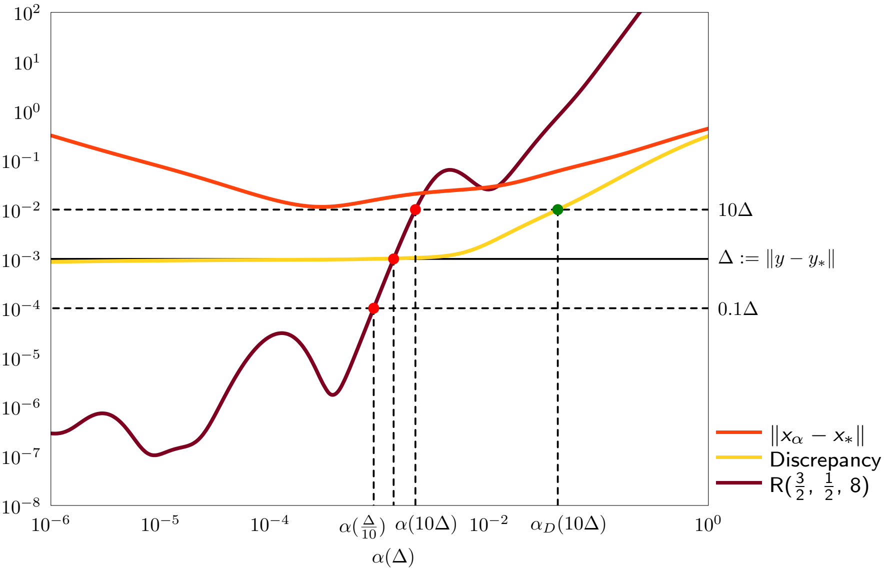
where $x_\alpha^+$ is the approximate solution with exact right-hand side and $\gamma_* = 1/2$ for Tikhonov method, $\gamma_* = m$ for iterated and extrapolated variants of Tikhonov method.

- Largest solution $\Rightarrow$ stability
- Smallest solution $\Rightarrow$ quasi-optimality
- If the solution is unique, quasi-optimality also holds for the largest solution. In most of our numerical experiments the solution was unique.

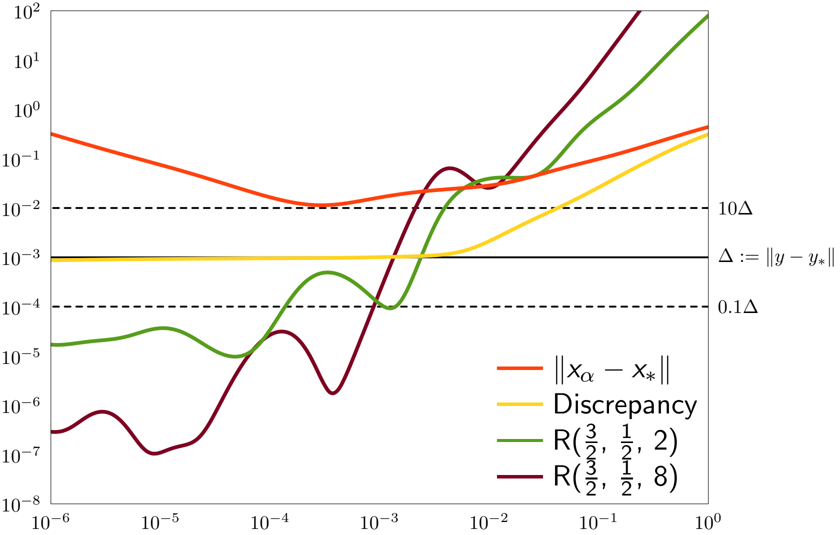
In the following we choose the largest solution.

The following 3 slides show the behaviour of functions $d(\alpha)$ in the problem 'phillips' from Hansen's Regularization Tools.

# Stability of choice $\alpha = \alpha(\delta)$ from rule $d(\alpha) = \delta$



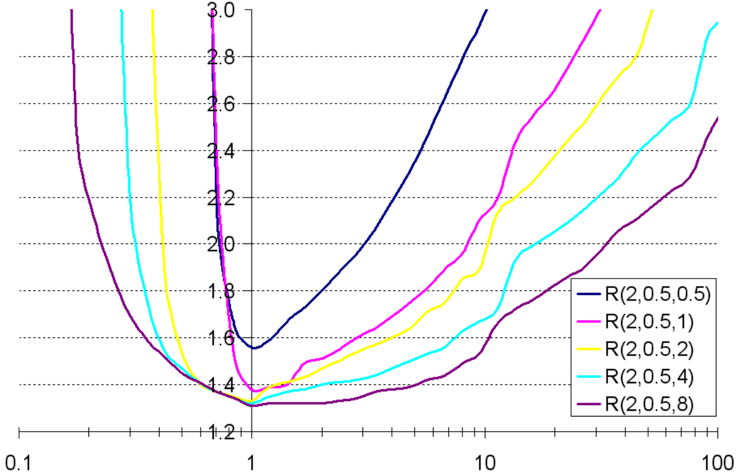
# Behavior of functions $d(\alpha)$ in rules $d(\alpha) = \delta$



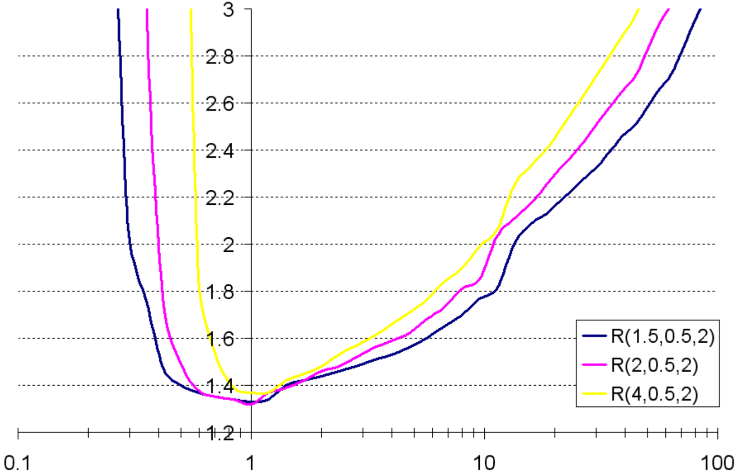
# Perturbed data and presentation of numerical results

- Numerical experiments on a large set of test problems (to be precisized on a later slide).
- For perturbed data we took $y = y_* + \Delta$, $\|\Delta\| = 0.3, 10^{-1}, \ldots, 10^{-6}$ with 10 different normally distributed perturbations $\Delta$ generated by computer.
- Problems were solved by Tikhonov method, assuming that the noise level is $\delta = \rho\|y - y_*\|$. Thus $\rho > 1$ corresponds to overestimation of the true noise level, $\rho < 1$ to underestimation.
- To compare the rules, we present averages (over problems, perturbations $\Delta$ and runs) of error ratios $\|x_\alpha - x_*\|/e_{\text{opt}}$ as the function of the argument $\rho$, where $e_{\text{opt}}$ is minimal error in Tikhonov method.

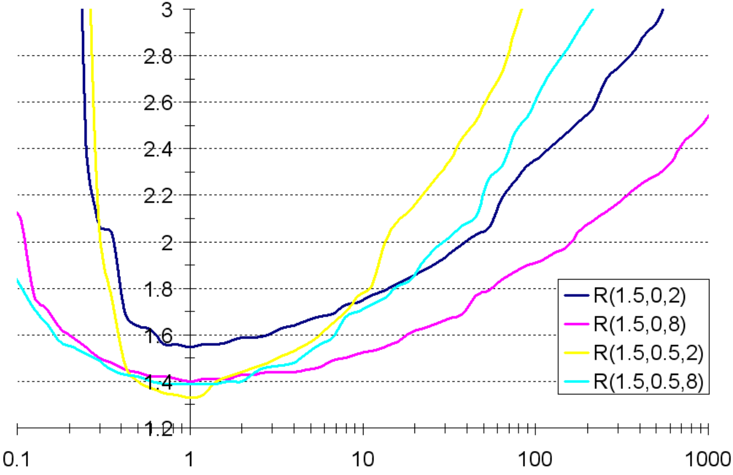
# Stability of rule R(q, l, k) increases if k increases



# Stability of rule R($q$, $l$, $k$) increases if $q$ decreases



# $l = 0.5$ is recommended ($l = 0$ is good if $\delta \gg \|y - y_*\|$)



# Case 2: heuristic rules not using noise level information

- **QC-rule for T-method** (analog of the quasioptimality criterion Q): Make the computations using the sequence of parameters $\alpha_i = r^{-i}$, $i = 0, 1, \ldots, r > 1$ (for example $r = 1.1$). Take $\alpha_{QC} = \alpha_i$ as the minimizer of the function $\psi(\alpha_i) = (1 + \alpha_i \|A\|^{-2}) \|x_{m,\alpha_i} - x_{m+1,\alpha_i}\|$ in the interval $[\max(m\sigma_{\min}, \underline{\alpha}), 1]$, where $\sigma_{\min}$ is the smallest eigenvalue of the discretized version of the operator $A^*A$ and $\underline{\alpha}$ is the largest $\alpha_i$, for which the value of $\psi(\alpha_i)$ is $C = 5$ times larger than its value at its current minimum.
- **QC-rule for Landweber method**: compute $\psi(n) := \|x_n - x_{2n+100}\|$ for $n = 1, 2, \ldots$ and take $n = n_{QC}$ as the minimizer of the function $\psi(n)$ for $n \in [1, N]$, where $N$ is the smallest $n$ for which the value of $\psi(n)$ is $C = 20$ times larger than its value at its current minimum.
- L-curve rule, GCV-rule, Hanke-Raus rule and Brezinski-Rodriguez-Seatzu rule gave in our numerical experiments not so good results as rules Q and QC, especially in case of smooth solutions.

# Heuristic rule for Lavrentiev method

We propose to find the parameter $\alpha$ as the minimizer of the function

$$\bar{g}_k(\alpha) = \alpha^{-2} \sum_{j=0}^{2k} c_j \| D_\alpha^{j/2} B_\alpha (Ax_\alpha - y) \|^2, \quad c_j = j/3 + 1,\ j = 0,\ 1, \ldots, 2k.$$

We propose to make computations on the sequence of parameters $\alpha_i = r^{-i}$, $i = 0, 1, \ldots,$ $r = 1.1$. The parameter $\alpha_i$ is found as the minimizer of the function $\bar{g}_k(\alpha)$ in the interval $[\underline{\alpha}, 1]$, where $\underline{\alpha}$ is the largest $\alpha_i$ for which the value of $\bar{g}_k(\alpha_i)$ is $C$ times larger than its value at its current minimum. We used the value $C = 1.2$.

# Hansen's test problems used in numerical tests.

Set I of test problems, P. C. Hansen's *Regularization Tools*.

| Nr | Problem | $\text{cond}_{100}$ | selfadj | Description |
|---|---|---|---|---|
| 1 | baart | 5e+17 | no | (Artificial) Fredholm integral equation of the first kind |
| 2 | deriv2 | 1e+4 | yes | Computation of the second derivative |
| 3 | foxgood | 1e+19 | yes | A problem that does not satisfy the discrete Picard condition |
| 4 | gravity | 3e+19 | yes | A gravity surveying problem |
| 5 | heat | 2e+38 | no | Inverse heat equation |
| 6 | ilaplace | 9e+32 | no | Inverse Laplace transform |
| 7 | phillips | 2e+6 | yes | An example problem by Phillips |
| 8 | shaw | 5e+18 | yes | An image reconstruction problem |
| 9 | spikes | 3e+19 | no | Test problem whose solution is a pulse train of spikes |
| 10 | wing | 1e+20 | no | Fredholm integral equation with discontinuous solution |

# Brezinski-Rodriguez-Seatzu problems

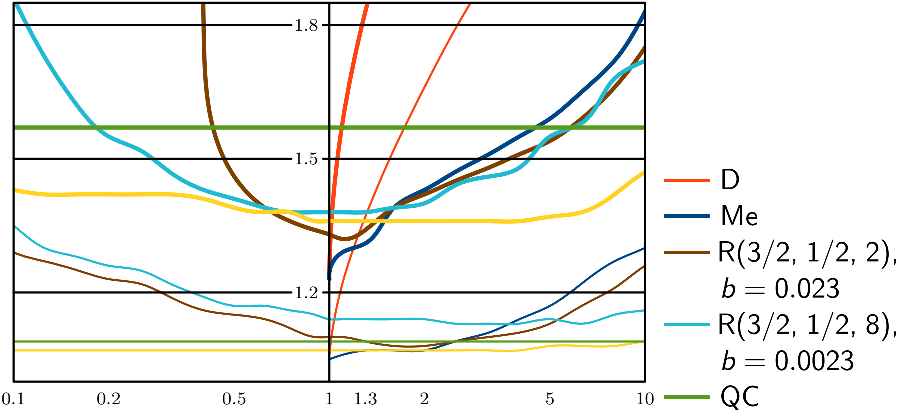
Set II of test problems, *Numerical Algorithms* 2008, 49, 1–4, pp 85–104.

| Nr | Problem | $\text{cond}_{100}$ | selfadj | Description |
|---|---|---|---|---|
| 11 | gauss | 6e+18 | yes | Test problem with Gauss matrix $a_{ij} = \sqrt{\frac{\pi}{2\sigma}} e^{-\frac{\sigma}{2(i-j)^2}}$, where $\sigma = 0.01$ |
| 12 | hilbert | 4e+19 | yes | Test problem with Hilbert matrix $a_{ij} = \frac{1}{i+j-1}$ |
| 13 | lotkin | 2e+21 | no | Test problem with Lotkin matrix (same as Hilbert matrix, except $a_{1j} = 1$) |
| 14 | moler | 2e+4 | yes | Test problem with Moler matrix $A = B^T B$, where $b_{ii} = 1$, $b_{i,i+1} = 1$, and $b_{ij} = 0$ otherwise |
| 15 | pascal | 1e+60 | yes | Test problem with Pascal matrix $a_{ij} = \binom{i+j-2}{i-1}$ |
| 16 | prolate | 1e+17 | yes | Test problem with a symmetric, ill-conditioned Toeplitz matrix |

# Solution vectors for BRS-problems

| Description | $\overline{x}_i$ |
|---|---|
| constant | $1$ |
| linear | $\frac{i}{N}$ |
| quadratic | $\left(\frac{i-\lfloor\frac{N}{2}\rfloor}{\lceil\frac{N}{2}\rceil}\right)^2$ |
| sinusoidal | $\sin\frac{2\pi(i-1)}{N}$ |
| linear+sinusoidal | $\frac{i}{N}+\frac{1}{4}\sin\frac{2\pi(i-1)}{N}$ |
| step function | $\begin{cases}0, & \text{if } i\leq\lfloor\frac{N}{2}\rfloor\\ 1, & \text{if } i>\lfloor\frac{N}{2}\rfloor\end{cases}$ |

# Averages (thick lines) and medians (thin lines) of error ratios in various rules in dependence of $\rho = \delta / \|y - y_*\|$



# Preferences of rules, in dependence of the accuracy of noise level information $\rho = \delta / \|y - y_*\|$

- T-method
  - If we are sure that $\rho \in [1, 2]$, then we recommend the rule Me.
  - In case $\rho \in [0.6, 2]$ we recommend the rule R(3/2, 1/2, 2), $b = 0.023$.
  - For even less information about the noise level, we recommend the rule QC.
- Landweber method
  - If we are sure that $\rho \in [1, 1.1]$, then we recommend MEe-rule.
  - Otherwise we recommend the QC-rule.

# Conclusions

- We propose for Tikhonov method and its modifications and for the methods of Lavrentiev and Landweber a family of rules R($q$, $l$, $k$) for approximate noise level, where $\underline{q} \leq q < \infty$, $l \geq 0$, $k \geq l/q$, $2k$, $2l \in \mathbb{N}$, $(m+1)q \in \mathbb{N}$ for T-method, $3q \in \mathbb{N}$ for Lavrentiev method.
- If $k > l/q$ and $\frac{\|y - y_*\|}{\delta} \leq C = \text{const}$ as $\delta \to 0$, then we have $\|x_\alpha - x_*\| \to 0$ $(\delta \to 0)$.
- Certain rules from the family gave in numerical experiments good results in case of several times over- or underestimated noise level.

# Bibliography


1. U. Hämarik, R. Palm, T. Raus. On minimization strategies for choice of the regularization parameter in ill-posed problems. *Numerical Functional Analysis and Optimization*, 30(9&10):924–950, 2009.
2. R. Palm. *Numerical comparison of regularization algorithms for solving ill-posed problems*. PhD thesis, University of Tartu, 2010. `http://hdl.handle.net/10062/14623`.
3. U. Hämarik, R. Palm, T. Raus. A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level. *Journal of Computational and Applied Mathematics*, 236(8):2146–2157, 2012.
4. U. Hämarik, R. Palm, T. Raus. A family of rules for the choice of the regularization parameter in Lavrentiev method in the case of rough estimate of the noise level of the data. *Journal of Inverse and Ill-Posed Problems*, 20(5-6), 831 - 854, 2012.
5. U. Hämarik, U. Kangro, R. Palm, T. Raus, U. Tautenhahn. Monotonicity of error of regularized solution and its use for parameter choice. Submitted to *Inverse Problems in Science and Engineering*

18th International Conference Mathematical Modelling and Analysis (MMA2013)
Fourth International Conference Approximation Methods and Orthogonal Expansions (AMOE2013)
May 27 - 30, 2013, Tartu, Estonia (http://www.ut.ee/mma-amoe2013/)
The conference is dedicated to the 75th birthday of professor Gennadi Vainikko
Confirmed plenary speakers:

- Hermann Brunner (Hong Kong, China)
- Raimondas Ciegis (Lithuania)
- Zdzislaw Jackiewicz (USA)
- Barbara Kaltenbacher (Austria)
- Rainer Kress (Germany)
- M. Zuhair Nashed (USA)
- Helmut Neunzert (Germany)
- Sergei Pereverzyev (Austria)
- Ian H. Sloan (Australia)
- Gennadi Vainikko (Estonia)