Heuristic parameter choice in Tikhonov method from minimizers of the quasi-optimality function

Toomas Raus and Uno Hämarik

University of Tartu, Estonia

31.10.2017, Rio de Janeiro

Contents

Problem

$$Au = f_*, \qquad A \in \mathcal{L}(H,F), \qquad f_* \in \mathcal{R}(A), \tag{1}$$

f_*, f - exact and noisy data. Range $\mathcal{R}(A)$ non-closed, kernel $\mathcal{N}(A)$ non-trivial. Tikhonov method, using f_*, f:

$$u_\alpha^+ = (\alpha I + A^*A)^{-1} A^* f_*, \qquad u_\alpha = (\alpha I + A^*A)^{-1} A^* f$$

Problem: how to choose the regularization parameter $\alpha > 0$?
Denote

$$e_1(\alpha) := \left\| u_\alpha^+ - u_* \right\| + \left\| u_\alpha - u_\alpha^+ \right\| . \tag{2}$$

The aim: find rule R, choosing the parameter α_R with the property ('pseudooptimality' property)

$$\left\| u_{\alpha_R} - u_* \right\| \leq \text{const} \quad \min_{\alpha > 0} e_1(\alpha)$$

Quasioptimality criterion

Quasioptimality criterion:

$$\alpha_Q := \text{argmin}_{\alpha>0}\psi_Q(\alpha),$$

where

$$\psi_Q(\alpha) := \alpha \left\| \frac{du_\alpha}{d\alpha} \right\| = \alpha^{-1} \left\| A^*(Au_{2,\alpha} - f) \right\|,$$
$$u_{2,\alpha} = (\alpha I + A^*A)^{-1}(\alpha u_\alpha + A^*f).$$

We search the regularization parameter from the set

$$\Omega = \{\alpha_j : \alpha_j = q\alpha_{j-1}, j = 1, 2, ..., M, 0 < q < 1\},$$

where α_0, M, q are given.
We show that at least one of local minimizers from the set Ω is psudooptimal and we show how to find it.

Rules for the choice of α, if $\delta \geq \|f - f_*\|$ is known

1) Discrepancy principle :

$$b_1\delta \leq \|Au_\alpha - f\| \leq b_2\delta, \quad b_1 \geq 1.$$

2) Modified discrepancy principle :

$$(Au_\alpha - f, Au_{2,\alpha} - f)^{1/2} = \delta, \quad u_{2,\alpha} = (\alpha I + A^*A)^{-1}(\alpha u_\alpha + A^*f).$$

3) Monotone error rule (ME-rule)

$$\frac{(Au_\alpha - f, Au_{2,\alpha} - f)}{\|Au_{2,\alpha} - f\|} = \delta$$

gives $\alpha_{\mathrm{ME}} \geq \alpha_{opt} := \operatorname{argmin}\|u_\alpha - u_*\|$. We recommend $\alpha_{MEe} := 0.4\alpha_{ME}$; typically $\|u_{\alpha_{\mathrm{MEe}}} - u_*\|/\|u_{\alpha_{\mathrm{ME}}} - u_*\| \in (0.7, 0.9)$.

Rules for the choice of α, if $\delta \geq \|f - f_*\|$ is unknown

1) The quasi-optimality criterion: $\alpha = \alpha_Q$ is the global minimizer of the function

$$\psi_Q(\alpha) = \alpha \left\| \frac{du_\alpha}{d\alpha} \right\| = \alpha^{-1} \left\| A^* \left(Au_{2,\alpha} - f \right) \right\|. \qquad (3)$$

2) The Hanke-Raus rule: $\alpha = \alpha_{HR}$ is the global minimizer of the function

$$\psi_{HR}(\alpha) = \alpha^{-1/2}(A_\alpha - f, Au_{2,\alpha} - f)^{1/2}.$$

3) L-curve rule: on graph with log-log scale, on x-axis $\|Au_\alpha - f\|$ and on y-axis $\|u_\alpha\|$ the corner point is used.
4) Reginska's rule: global minimum point of the function

$$\psi_{RE}(\alpha) = \|Au_\alpha - f\| \, \|u_\alpha\|^\tau, \tau \geq 1.$$

5) HME-rule: $\alpha = \alpha_{HME}$ is chosen as the global minimizer of the function

$$\psi_{HME}(\alpha) = \alpha^{-1/2} \frac{(Au_\alpha - f, Au_{2,\alpha} - f)}{\|Au_{2,\alpha} - f\|}.$$

Parameters set, regularization need

We use the set of parameters

$$\Omega = \{\alpha_j : \alpha_j = q\alpha_{j-1}, \quad j = 1, 2, ..., M, \quad 0 < q < 1\}, \tag{4}$$

where α_0, q, α_M are given. We search local minimizer of $\psi_Q(\alpha)$ in the interval $[\max(\alpha_M, \lambda_{min}), \alpha_0]$, where λ_{min} is the minimal eigenvalue of the matrix $A^T A$ of the discretized problem.
We say that the discretized problem $Au = f$ do not need regularization if

$$e_1(\lambda_{min}) = \min_{\alpha\in\Omega, \alpha\geq\lambda_{min}} e_1(\alpha).$$

If $\lambda_{min} > \alpha_M$ and the discretized problem do not need regularization then α_M is the proper parameter while then it is easy to show the error estimate

$$\|u_{\alpha_M} - u_*\| \leq e_1(\alpha_M) \leq 2 \min_{\alpha\in\Omega} e_1(\alpha).$$

Searching the parameter from the interval $[\max(\alpha_M, \lambda_{min}), \alpha_0]$ means the a priori assumption that the discretized problem needs regularization.

Test problems

Hansen's 10 test problems: Baart, Deriv2,Foxgood,Gravity,Heat,Ilaplace, Phillips,Shaw,Spikes,Wing.

The problems are scaled in such a way that the 2-norms of A and f are 1.
The dicretization number $n = 100$.
Relative noise levels: $\delta_{rel} := \|f - f_*\| / \|f_*\| = 10^{-j}, j = 1, 2, ..., 6$.
Noisy right-hand side: $f = f_* + \delta_{rel} \|f_*\| \|\xi\|^{-1} \xi$, where components ξ_i have standard normal distribution. For each noise level we considered 20 runs.

Sequence of the parameters: $\alpha_0 = 1, \alpha_M = 10^{-18}, q = 0.95$.
case $p = 0$. Original problems.
case $p = 2$. Problems, where the exact solution u_* is replaced by $A^T A u_*$.

Let λ_{min} be the minimal eigenvalue of the matrix $A^T A$.
Tables below show for different rules R the error ratios

$$E = \frac{\|u_{\alpha_R} - u_*\|}{\|u_{\alpha_*} - u_*\|} = \frac{\|u_{\alpha_R} - u_*\|}{\min_{\alpha \in \Omega} \|u_\alpha - u_*\|}.$$

Comparison of heuristic rules

Table: Averages of error ratios E and failure % (in parenthesis) for heuristic rules, $p = 0$

Problem	Λ	Quasiopt.	HR	HME	Reginska
Baart	1666	1.54	2.58	2.52	1.32
Deriv2	16	1.08	2.07	1.72	35.19 (3.3)
Foxgood	210	1.57	8.36	7.71	36.94 (10.8)
Gravity	4	1.13	2.66	2.32	20.49 (0.8)
Heat	$4 * 10^{29}$	> 100 (66.7)	1.64	1.48	23.40 (4.2)
Ilaplace	16	1.24	1.94	1.81	1.66
Phillips	9	1.09	2.27	1.91	> 100 (44.2)
Shaw	290	1.43	2.34	2.23	1.80
Spikes	1529	1.01	1.03	1.03	1.01
Wing	9219	1.40	1.51	1.51	1.18

$\Lambda = \max_{\lambda_k > \max(\alpha_M, \lambda_n)} \lambda_k / \lambda_{k+1}$ of consecutive eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ of the matrix $A^T A$ in the interval $[\max(\alpha_M, \lambda_n), 1]$.

Local minimum points of the function $\psi_Q(\alpha)$

Lemma 1

The function $\psi_Q(\alpha)$ has the estimate (see (2) for notation $e_1(\alpha)$)

$$\psi_Q(\alpha) \le e_1(\alpha) = \left\| u_\alpha^+ - u_* \right\| + \left\| u_\alpha - u_\alpha^+ \right\|. \tag{5}$$

Remark 1

Note that $\lim_{\alpha\to\infty} \psi_Q(\alpha) = 0$, *but* $\lim_{\alpha\to\infty} e_1(\alpha) = \|u_*\|$. *Therefore in the case of too large α_0 this α_0 may be global (or local) minimizer of the function $\psi_Q(\alpha)$. We recommend to take $\alpha_0 = c\,\|A^*A\|\,, c \le 1$ or to minimize the function $\tilde{\psi}_Q(\alpha) := (1 + \alpha/\,\|A^*A\|)\psi_Q(\alpha)$ instead of $\psi_Q(\alpha)$.*

Lemma 2

Denote $\psi_{QD}(\alpha) = (1-q)^{-1}\,\|u_\alpha - u_{q\alpha}\|$. Then it holds

$$\psi_Q(\alpha) \le \psi_{QD}(\alpha) \le q^{-1}\psi_Q(q\alpha).$$

Parameter $\alpha_k, 0 \leq k \leq M-1$ is the local minimum point of the sequence $\psi_Q(\alpha_k)$, if $\psi_Q(\alpha_k) < \psi_Q(\alpha_{k+1})$ and in case $k > 0$ there exists index $j \geq 1$ such, that

$$\psi_Q(\alpha_k) = \psi_Q(\alpha_{k-1}) = ... = \psi_Q(\alpha_{k-j+1}) < \psi_Q(\alpha_{k-j}).$$

Parameter α_M is the local minimum point if there $\exists j \geq 1$ so, that

$$\psi_Q(\alpha_M) = \psi_Q(\alpha_{M-1}) = ... = \psi_Q(\alpha_{M-j+1}) < \psi_Q(\alpha_{M-j}).$$

Denote the set of local minimum points:

$$L_{min} = \left\{\alpha_{min}^{(k)} : \alpha_{min}^{(1)} > \alpha_{min}^{(2)} > ... > \alpha_{min}^{(K)}\right\}.$$

Parameter α_k is the local maximum point of the sequence $\psi_Q(\alpha_k)$ if $\psi_Q(\alpha_k) > \psi_Q(\alpha_{k+1})$ and there exists index $j \geq 1$ so, that

$$\psi_Q(\alpha_k) = \psi_Q(\alpha_{k-1}) = ... = \psi_Q(\alpha_{k-j+1}) > \psi_Q(\alpha_{k-j}).$$

We denote by $\alpha_{max}^{(k)}$ the local maximum point between the local minimum points $\alpha_{min}^{(k+1)}$ and $\alpha_{min}^{(k)}, 1 \leq k \leq K-1$. Denote $\alpha_{max}^{(0)} = \alpha_0, \alpha_{max}^{(K)} = \alpha_M$. Then by the construction

$$\alpha_{max}^{(0)} \geq \alpha_{min}^{(1)} > \alpha_{max}^{(1)} > ... > \alpha_{max}^{(K-1)} > \alpha_{min}^{(K)} \geq \alpha_{max}^{(K)}.$$

Theorem 3

Local minimum points of the function $\psi_Q(\alpha)$ have estimates:

1. $$\min_{\alpha \in L_{min}} \|u_\alpha - u_*\| \le q^{-1} C \min_{\alpha_M \le \alpha \le \alpha_0} e_1(\alpha),$$

   where

   $$C := 1 + \max_{1 \le k \le K} \max_{\alpha_j \in \Omega, \alpha_{max}^{(k)} \le \alpha_j \le \alpha_{max}^{(k-1)}} T\left(\alpha_{min}^{(k)}, \alpha_j\right) \le 1 + c_q \ln\left(\frac{\alpha_0}{\alpha_M}\right),$$

   $$T(\alpha, \beta) := \frac{\|u_\alpha - u_\beta\|}{\psi_Q(\beta)}, \qquad c_q := \left(q^{-1} - 1\right) / \ln q^{-1} \to 1 \textit{ if } q \to 1.$$

2. *Let* $u_* = (A^*A)^{p/2} v$, $\|v\| \le \rho$, $0 < p \le 2$ *and* $\alpha_0 = 1$. *If* $\delta_0 := \sqrt{\alpha}_M \le \|f - f_*\|$, *then*

   $$\min_{\alpha \in L_{min}} \|u_\alpha - u_*\| \le c_p \ln \frac{\|f - f_*\|}{\delta_0} \rho^{\frac{1}{p+1}} \left|\ln \|f - f_*\|\right| \|f - f_*\|^{\frac{p}{p+1}}.$$

Table: Results for the set L_{min}, $p = 0$

Problem	ME	MEe	DP	Best of L_{min}		$\lvert L_{min} \rvert$		Apost. C	
	Aver E	Aver E	Aver E	Aver E	Max E	Aver	Max	Aver	Max
Baart	1.43	1.32	1.37	1.23	2.51	6.91	8	3.19	3.72
Deriv2	1.09	1.08	1.28	1.08	1.34	2.00	2	3.54	4.49
Foxgood	1.98	1.42	1.34	1.47	6.19	3.63	6	3.72	4.16
Gravity	1.40	1.13	1.16	1.13	1.83	1.64	3	3.71	4.15
Heat	1.19	1.03	1.05	1.12	2.36	3.19	5	3.92	4.50
Ilaplace	1.33	1.21	1.26	1.20	2.56	2.64	5	4.84	6.60
Phillips	1.27	1.02	1.02	1.06	1.72	2.14	3	3.99	4.66
Shaw	1.37	1.24	1.28	1.19	2.15	4.68	7	3.48	4.43
Spikes	1.01	1.00	1.01	1.00	1.02	8.83	10	3.27	3.70
Wing	1.16	1.13	1.15	1.09	1.38	5.20	6	3.07	3.72
Total	1.32	1.16	1.19	1.16	6.19	4.09	10	3.67	6.60

Restricted set of the local minimizers of the function $\psi_Q(\alpha)$

Two phases for restriction of the set L_{min}. In the first phase we remove from L_{min} local minimizers in interval, where the function $\phi(\alpha) = (Au_\alpha - f, Au_{2,\alpha} - f)^{1/2}$ decreases only a little bit. On the second phase we remove from set obtained on the first phase these local minimizers for which the function $\psi_Q(\alpha)$ for decreasing α-values has only small growth before the next decrease.

1. Denote $\delta_M := (Au_{\alpha_M} - f, Au_{2,\alpha_M} - f)^{1/2}$ and by $\alpha = \alpha_{MD}$ the parameter for which $(Au_\alpha - f, Au_{2,\alpha} - f)^{1/2} = b\delta_M, \quad b > 1$. Denote $\alpha_{MDQ} := \min(\alpha_{MD}, \alpha_Q)$, where $\alpha_Q \in L_{min}$ is the global minimizer of the function $\psi_Q(\alpha)$ on the set Ω. Let $\alpha_{max}^{(k_0)} \leq \alpha_{MDQ} < \alpha_{max}^{(k_0-1)}$ for some $k_0, 1 \leq k_0 \leq K$. Then we get the set $L_{min}^0 = \left\{\alpha_{min}^{(k)} : 1 \leq k \leq k_0\right\}$. In the case $\alpha_{max}^{(k0)} \leq \alpha_{MDQ} \leq \alpha_{min}^{(k0)}$ we change denotation to $\alpha_{max}^{(k0)} := \alpha_{min}^{(k0)}$.

2. We remove from the set L^0_{min} these local minimizers $\alpha^{(k)}_{min}$ and following maximizers $\alpha^{(k)}_{max}$, which satisfy the following conditions:

$$\alpha^{(k)}_{min} \neq \alpha^{(k)}_{max}; \qquad \frac{\psi_Q(\alpha^{(k)}_{max})}{\psi_Q(\alpha^{(k)}_{min})} \leq c_0; \qquad \frac{\psi_Q(\alpha^{(k)}_{min})}{\min_{j\leq k} \psi_Q(\alpha^{(j)}_{min})} \leq c_0,$$

where $c_0 > 1$ is some constant. We denote by

$$L^*_{min} := \left\{ \overline{\alpha}^{(k)}_{min} : \overline{\alpha}^{(1)}_{min} > \overline{\alpha}^{(2)}_{min} > ... > \overline{\alpha}^{(k_*)}_{min} \right\}$$

the set of minimizers remained in L^0_{min} and denote the remained maximizers by $\overline{\alpha}^{(k)}_{max} : \overline{\alpha}^{(0)}_{max} > \overline{\alpha}^{(1)}_{min} > ... > \overline{\alpha}^{(k_*)}_{max}$. According to this algorithm the following inequalities hold:

$$\overline{\alpha}^{(0)}_{max} \geq \overline{\alpha}^{(1)}_{min} > \overline{\alpha}^{(1)}_{max} > ... > \overline{\alpha}^{(k_*-1)}_{max} > \overline{\alpha}^{(k_*)}_{min} \geq \overline{\alpha}^{(k_*)}_{max}.$$

Theorem 4

The following estimates hold for the local minimum points of the set L^*_{min}:

$$\min_{\alpha \in L^*_{min}} \|u_\alpha - u_*\| \leq \max\left\{ q^{-1} C_1 \min_{\alpha_M \leq \alpha \leq \alpha_0} e_1(\alpha), C_2(b) \min_{\alpha_M \leq \alpha \leq \alpha_0} e_2(\alpha, \delta_*) \right\},$$

where

$$C_1 := 1 + \max_{1 \leq k \leq k_*} \max_{\alpha_j \in \Omega, \overline{\alpha}^{(k)}_{max} \leq \alpha_j \leq \overline{\alpha}^{(k-1)}_{max}} T\left(\overline{\alpha}^{(k)}_{min}, \alpha_j\right) \leq 1 + c_0 c_q \ln\left(\frac{\alpha_0}{\overline{\alpha}^{(k_*)}_{max}}\right)$$

and $\delta_* = \max(\delta_M, \|f - f_*\|)$, $C_2(b) = b + 2$.
Moreover, if $u_* = (A^*A)^{p/2} v$, $\|v\| \leq \rho$, $p > 0$, $\alpha_0 = 1$ *and* $\delta_0 := \sqrt{\alpha}_M \leq \|f - f_*\|$, *then*

$$\min_{\alpha \in L^*_{min}} \|u_\alpha - u_*\| \leq c_0 c_p \ln \frac{\|f - f_*\|}{\delta_0} \rho^{\frac{1}{p+1}} \left|\ln \|f - f_*\|\right| \|f - f_*\|^{\frac{p}{p+1}}, 0 < p \leq 2.$$

Table: Results about the set L^*_{min}, $p = 0$

Problem	Best of L^*_{min}		$\lvert L^*_{min}\rvert$		Apost. C_1		$\lvert L^*_{min}\rvert = 1$
	Aver E	Max E	Aver	Max	Aver	Max	%
Baart	1.40	2.91	1.41	3	6.38	7.93	60.8
Deriv2	1.08	1.34	2.00	2	3.54	4.49	100
Foxgood	1.57	6.69	1,00	1	4.39	4.92	100
Gravity	1.14	2.15	1.00	1	3.02	3.95	100
Heat	1.12	2.36	2.05	3	5.08	5.38	0
Ilaplace	1.23	2.56	1.00	1	4.68	6.68	100
Phillips	1.06	1.72	2.10	3	3.97	4.66	90.0
Shaw	1.39	3.11	1.16	2	5.89	8.06	84.2
Spikes	1.01	1.03	1.64	3	10.07	11.82	55.0
Wing	1.30	1.84	2.18	4	3.03	6.63	1.7
Total	1.23	6.69	1.55	4	5.01	11.82	69.2

Choice of the parameter from the set L^*_{min}

Now we give algorithm for choice of the regularization parameter from the set L^*_{min}.

1. If the set L^*_{min} contains only one parameter, we take this for the regularization parameter.
2. If the set L^*_{min} contains two parameters one of which is α_M, we take another parameter $\alpha \neq \alpha_M$.
3. If the set L^*_{min} contains after possible elimination of α_M more than one parameter, we may use for parameter choice the following algorithms.

a) Let α_Q, α_{HR} be global minimizers of the functions $\psi_Q(\alpha)$, $\psi_{HR}(\alpha)$ respectively on the interval $[\max(\alpha_M, \lambda_{min}), \alpha_0]$. Let $\alpha_{Q1} := \max(\alpha_Q, \alpha_{HR})$. Choose from the set L^*_{min} the largest parameter α, which is smaller or equal to α_{Q1}.

b) Let α_{RE} be the global minimizer of the function $\psi_{RE}(\alpha)$ on the interval $[\max(\alpha_M, \lambda_{min}), \alpha_0]$. Let α_{Q2} be the global minimizer of the function $\psi_Q(\alpha)$ on the interval $[\alpha_{RE}, \alpha_0]$. Choose from the set L^*_{min} the largest parameter α, which is smaller or equal to α_{Q2}.

c) For the parameters from L^*_{min} we compute value $R(\alpha) = \frac{\psi_{HR}(\alpha)}{\|u_\alpha\|}$ which we consider as the rough estimate for the relative error $\frac{\|u_\alpha - u_*\|}{\|u_*\|}$ under assumption that parameter α is near to the optimal parameter. We choose for the regularization parameter the smallest parameter α_* from the set L^*_{min}, which satisfies the condition $R(\alpha_*) \leq C^* \min_{\alpha \in L^*_{min}, \alpha > \alpha_*} R(\alpha)$. We recommend to choose the constant C^* from the interval $5 \leq C^* \leq 10$. In the numerical experiments we used $C^* = 5$.

Table: Averages and maximums of error ratios E in case of different heuristic algorithms, $p = 0$

Problem	Algorithm a)		Algorithm b)		Algorithm c)	
	Aver E	Max E	Aver E	Max E	Aver E	Max E
Baart	1.83	3.63	1.61	2.91	1.61	2.91
Deriv2	1.08	1.34	1.08	1.34	1.08	1.34
Foxgood	1,57	6.69	1.57	6.69	1.57	6.69
Gravity	1.14	2.15	1.14	2.15	1.14	2.15
Heat	1.12	2.36	1.12	2.36	1.12	2.36
Ilaplace	1.23	2.56	1.23	2.56	1.23	2.56
Phillips	1.06	1.72	1.06	1.72	1.06	1.72
Shaw	1.48	3.64	1.45	3.64	1.45	3.64
Spikes	1.01	1.03	1.01	1.03	1.01	1.03
Wing	1.50	1.86	1.38	2.04	1.32	1.84
Total	1.30	6.69	1.26	6.69	1.26	6.69

Table: Results of the numerical experiments, $p = 2$

Problem	ME Aver E	MEe Aver E	Best of L_{min} Aver E	$\lvert L_{min} \rvert$ Aver	Best of L^*_{min} Aver E	$\lvert L^*_{min} \rvert$ Aver	$\lvert L^*_{min} \rvert = 1$ %
Baart	1.86	1.19	1.18	4.74	1.41	1.02	98.3
Deriv2	1.10	1.19	1.03	2.00	1.03	2.00	100
Foxgood	1.56	1.13	1.14	2.08	1.20	1.00	100
Gravity	1.33	1.05	1.09	1.72	1.11	1.00	100
Heat	1.13	1.12	1.05	2.10	1.05	2.10	0
Ilaplace	1.47	1.06	1.11	2.73	1.11	1.00	100
Phillips	1.26	1.06	1.04	2.10	1.04	2.10	90
Shaw	1.37	1.06	1.11	3.72	1.22	1.01	99.2
Spikes	1.85	1.12	1.19	4.78	1.31	1.00	100
Wing	1.67	1.14	1.22	4.53	1.73	1.01	99.2
Total	1.46	1.11	1.12	3.05	1.22	1.32	88.7