Convergence of heuristic parameter choice rules

Stefan Kindermann
Industrial Mathematics Institute
University of Linz, Austria

joint work with A. Neubauer

Regularization and Parameter Choices

Abstract linear ill-posed problems:

$$Tx = y_\delta \tag{1}$$

$T : X \to Y$, **Linear** Operator between **Hilbert spaces** X, Y.
Linear regularization by spectral filter functions

$$x_{\alpha,\delta} = R_\alpha y_\delta = g_\alpha(T^*T)T^*y_\delta \tag{2}$$

Notation:

- $x^\dagger$: exact (minimum norm) solution (unknown)
- $y = Tx^\dagger$ exact data
- $y_\delta = y + e_\delta$ noisy data
- $\delta = \|y - y_\delta\|$ Noise level (unknown)
- $x_{\alpha,\delta}$ regularized solution with noisy data
- x_α regularized solution with exact data (unknown)
- $r_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda)$

Framework:
Ill-posed case, deterministic noise, T injective (for simplification)

Parameter Choice

How to choose α ?

Bakushinskii Veto

If problem is ill-posed, then the worst-case error tends to 0

$$\sup_{\|y-y_\delta\|\leq\delta} \|x_{\alpha,\delta} - x^\dagger\| \to 0 \quad \text{as } \delta \to 0$$

only if $\alpha = f(\delta)$

Heuristic Parameter Choices:
Choose α independent from δ (but depending on y_δ).

Convergence analysis not in the worst case, but for reasonable 'nice' situations

Condition for convergence analysis
Noise $y - y_\delta$ has to be reasonably different from elements in $R(T)$

Minimization based parameter choice rules

Let $\psi : [0, \alpha_0] \times Y \to \mathbb{R}^+$ be a lower semicontinuous functional:

Minimization based parameter choice rules: Given y_δ

$$\alpha^*_{y_\delta} := \underset{0 \le \alpha \le \alpha_0}{\operatorname{argmin}}, \psi(\alpha, y_\delta)$$

Minimization based parameter choice rules: Examples

Quasi-Optimality Rule (Tikhonov; Glasko, Kriskin, Leonov; Neubauer)

$$\psi_{QO}(\alpha, y_\delta) = \int \lambda r_\alpha(\lambda) g_\alpha(\lambda) dF_\lambda \|y_\delta\|^2 \tag{3}$$

Hanke-Raus-Rules (Hanke, Raus)

$$\psi_{HR}(\alpha, y_\delta) = \frac{1}{\alpha} \int r_\alpha(\lambda)^{2+\frac{1}{\mu_0}} dF_\lambda \|y_\delta\|^2 \tag{4}$$

Special Case: $\mu_0 = \infty$ (Hanke, Raus ?)

$$\psi_{WR}(\alpha, y_\delta) = \frac{1}{\alpha} \| T x_{\alpha,\delta} - y_\delta\|^2 \tag{5}$$

L-Curve and modified L-Curve (Hansen, Reginska)

$$\psi_{mL}(\alpha, y_\delta) = \left(\int \lambda r_\alpha(\lambda)^2 dF_\lambda \|y_\delta\|^2\right)^{\frac{1}{2}} \left(\int \lambda g_\alpha(\lambda)^2 dF_\lambda \|y_\delta\|^2\right)^{\frac{\mu}{2}} \tag{6}$$

Minimization based parameter choice rules for Tikhonov regularization

For Tikhonov regularization:

$$\psi_{QO}(\alpha, y_\delta) = \int \frac{\alpha^2 \lambda}{(\alpha+\lambda)^4} dF_\lambda \|y_\delta\|^2$$

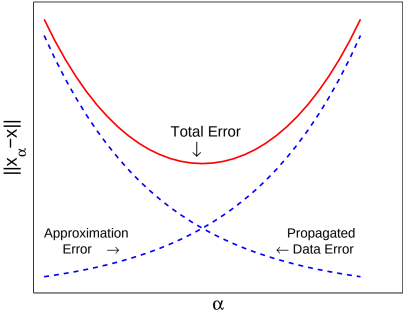
$$\psi_{HR}(\alpha, y_\delta) = \int \frac{\alpha^2}{(\alpha+\lambda)^3} dF_\lambda \|y_\delta\|^2$$

$$\psi_{WR}(\alpha, y_\delta) = \int \frac{\alpha}{(\alpha+\lambda)^2} dF_\lambda \|y_\delta\|^2$$

$$\psi_{ML} = \left(\int \frac{\lambda}{(\alpha+\lambda)^2} dF_\lambda \|y_\delta\|^2\right)^{\frac{\mu}{2}} \left(\int \frac{\alpha^2}{(\alpha+\lambda)^2} dF_\lambda \|y_\delta\|^2\right)^{\frac{1}{2}}$$

General analysis principle

Total error $=$ approximation error $+$ propagated data error



Successful parameter choice rules

$$\psi(\alpha, y_\delta) \sim \|x_{\alpha,\delta} - x_\alpha\| + \|x_\alpha - x^\dagger\|$$

Difficulties:
If $y_\delta \in R(T)$, then $\alpha = 0$
If $y_\delta \notin R(T)$ then $\alpha > 0$

Analysis: Standard Assumptions on ψ

Standard Assumptions on ψ:

- $$\psi : (0, \alpha_0] \times Y \to \mathbb{R} \text{ lower semicontinuous, with } \psi \geq 0 \quad (7)$$
- $$\psi(\alpha, y_\delta) > 0 \quad \forall \alpha \in (0, \alpha_0], y_\delta \neq 0 \quad (8)$$
- $$\psi(\alpha, y_\delta) \leq C \max\{\psi(\alpha, y_\delta - y), \psi(\alpha, y)\} \quad (9)$$
- For any sequence $(e_k) \in Y$ with $\lim_{k\to\infty} e_k = 0$ there is a sequence $\alpha_k \in (0, \alpha_0]$ with $\lim_{k\to\infty} \alpha_k = 0$ such that
$$\lim_{k\to\infty} \psi(\alpha_k, e_k) = 0 \quad (10)$$
- If $y \in D(T^\dagger)$
$$\lim_{\alpha\to 0} \psi(\alpha, y) = 0 \quad (11)$$

Analysis: Other Assumptions on ψ

Noise Conditions

There exists a set $\mathcal{N} \subset Y$ such that for all $e \in \mathcal{N}$

$$\lim_{\alpha \to 0} \psi(\alpha, e) = \infty \tag{12}$$

and for all $e \in \mathcal{N}$ with $\|e\| = \delta$ and $y_\delta = y + e$

$$\lim_{\delta \to 0} \psi(\alpha^*_{y_\delta}, e) = 0 \Rightarrow \|x_{\alpha^*_{y_\delta}, \delta} - x_{\alpha^*_{y_\delta}}\| \to 0 \tag{13}$$

Analysis: Other Assumptions on ψ

Noise Conditions

There exists a set $\mathcal{N} \subset Y$ such that for all $e \in \mathcal{N}$

$$\lim_{\alpha \to 0} \psi(\alpha, e) = \infty \qquad (12)$$

and for all $e \in \mathcal{N}$ with $\|e\| = \delta$ and $y_\delta = y + e$

$$\lim_{\delta \to 0} \psi(\alpha^*_{y_\delta}, e) = 0 \Rightarrow \|x_{\alpha^*_{y_\delta}, \delta} - x_{\alpha^*_{y_\delta}}\| \to 0 \qquad (13)$$

Noise condition is satisfied if for $y - y_\delta \in \mathcal{N}$

$$f(\|x_{\alpha,\delta} - x_\alpha\|) \leq \psi(\alpha, y - y_\delta)$$

Convergence Theorem

Theorem

Let the standard assumptions and noise conditions hold, then

$$\alpha^*_{y_\delta} > 0$$

$$\lim_{\delta \to 0} \psi(\alpha^*_{y_\delta}, y_\delta) = 0$$

$$x_{\alpha^*_{y_\delta}} \to x^\dagger \quad \textit{for } \delta \to 0$$

Tikhonov regularization, quasioptimality: [Glasko, Kriskin]; Banach spaces: [Lorenz, Jin]

Convergence Rate Theorem

Theorem

Let the standard assumption hold and let $x^\dagger = (A^*A)^\nu \omega$ *satisfy as Hölder source condition Let* f, g, k *be positive continuous monotonically increasing functions mapping* $[0, \infty]$ *onto itself. Moreover let* ψ *satisfy the following conditions:*

$$f(\|x_\alpha - x_{\alpha,\delta}\|) \leq \quad \psi(\alpha, y_\delta - y) \quad \leq f(\frac{\delta}{\sqrt{\alpha}}) \qquad \forall y_\delta - y \in \mathcal{N}, \alpha$$

$$\psi(\alpha, y) \quad \leq g_{x^\dagger}(\alpha) \qquad \forall y \in R(A)$$

$$k(\alpha) \leq \quad \psi(\alpha, y_\delta) \qquad y_\delta - y \in \mathcal{N}, \alpha, \delta \text{ suff. small}$$

With $r(\delta) = \inf f(\frac{\delta}{\sqrt{\alpha}}) + g_{x^\dagger}(\alpha)$ *we get the abstract convergence rate*

$$\|x_{\alpha,\delta} - x^\dagger\| \leq C \max\{f^{-1} \circ (r(\delta)), f^{-1} \circ (g_{x^\dagger} \circ k^{-1} \circ r(\delta))\} + h \circ k^{-1} \circ r(\delta)$$

Oracle Estimates

Theorem

Let ψ satisfy the standard assumptions and additionally let there exists a f as before such that

$$\begin{aligned} f(\|x_{\alpha,\delta} - x_\alpha\|) \leq \quad & \psi(\alpha, y_\delta - y) & \leq Cf(\|x_{\alpha,\delta} - x_\alpha\|) \quad \forall y - y_\delta \in \mathcal{N} \\ f(\|x_\alpha - x^\dagger\|) \leq \quad & \psi(\alpha, y) & \leq Cf(\|x_\alpha - x^\dagger\|). \end{aligned}$$

Moreover, let the regularization be monotonous in α. Then there exists a constant C independent of α, δ such that

$$\|x_{\alpha^*_{y_\delta},\delta} - x^\dagger\| \leq C\left(\inf_\alpha \|x_\alpha - x_{\alpha,\delta}\| + \|x_\alpha - x^\dagger\|\right) \tag{14}$$

Standard Assumptions

Let $r_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda) \neq 0$ for all $\lambda \in (0, \|T\|)$ Then the Standard Assumptions are satisfied for $\psi_{QO}, \psi_{HR}, \psi_{WR}$.
For ψ_{mL} all but the triangle inequality are satisfied.

Regular Filter Function

We consider **regular filter functions** in the following sense:

- $$|\lambda g_\alpha(\lambda)| \leq C \quad \lim_{\alpha \to 0} g_\alpha(\lambda) = \frac{1}{\lambda} \tag{15}$$
- $g_\alpha(\lambda)$ is monotonically decreasing with respect to α
- $g_\alpha(\lambda), r_\alpha(\lambda)$ are monotonically decreasing with respect to λ
- $|G_\alpha| \leq C\alpha^{-1}$
- For any $\gamma > 0$ there are constants c_1, c_2 such that
$$c_1 \leq r_\alpha(\gamma\alpha) \leq c_2 \quad \forall 0 < \alpha \leq \gamma^{-1}\eta$$
where η is a positive constant independent of γ
- there is a qualification, if $\mu_0 \leq \infty$, then γ_0, c_0 exists with $\lambda^{\mu_0}|r_\alpha(\lambda)| \geq c_0\alpha^{\mu_0}$ for all $\gamma_0\alpha \leq \lambda$

Assumptions are satisfied for most of the standard regularization methods (Tikhonov, Landweber)

Noise Conditions

$$t^2 \int_t^\infty \lambda^{-1} d\|F_\lambda(y - y_\delta)\|^2 \leq c \int_0^t \lambda d\|F_\lambda(y - y_\delta)\|^2 \quad \forall 0 < t \leq t_1 \tag{16}$$

We define the set $\mathcal{N}_1$ as

$$\mathcal{N}_1 := \{y - y_\delta \in Y \mid y - y_\delta \text{ satisfy}(16)\}$$

$$t \int_t^\infty \lambda^{-1} d\|F_\lambda(y - y_\delta)\|^2 \leq c \int_0^t d\|F_\lambda(y - y_\delta)\|^2 \quad \forall 0 < t \leq t_1 \tag{17}$$

We define the set $\mathcal{N}_2$ as

$$\mathcal{N}_2 := \{y - y_\delta \in Y \mid y - y_\delta \text{ satisfy}(17)\}$$

Note: $\mathcal{N}_1 \subset \mathcal{N}_2$

Convergence Results

Let g_α be a regular filter function
If $y - y_\delta \in \mathcal{N}_1$

$$\|x_{\alpha,\delta} - x_\alpha\|^2 \leq C\psi_{QO}(\alpha, y - y_\delta) \tag{18}$$

If $y - y_\delta \in \mathcal{N}_2$ then

$$\|x_{\alpha,\delta} - x_\alpha\|^2 \leq C\psi_{HR}(\alpha, y - y_\delta)$$

$$\|x_{\alpha,\delta} - x_\alpha\|^2 \leq C\psi_{WR}(\alpha, y - y_\delta)$$

In particular if $y - y_\delta \in \mathcal{N}_1$ then $\psi_{QO}, \psi_{HR}, \psi_{WR}$ yield convergence
In particular if $y - y_\delta \in \mathcal{N}_2$ then ψ_{HR}, ψ_{WR} yield convergence

Convergence Results: L-curve

For the L-curve method it can be shown that $\alpha^*_{y_\delta} > 0$ for some $y - y_\delta \notin R(T)$

Lower estimates are impossible:
Negative Result:
There cannot be a set $\mathcal{N} \notin R(T)$ such that for $p > \mu$ an estimate

$$\|x_{\alpha,\delta} - x_\alpha\|^p \leq C\psi_{ML}(\alpha, y - y_\delta) \quad \forall y - y_\delta \in \mathcal{N} \qquad (19)$$

holds.

(our) Convergence theory not applicable to L-curve

Convergence Rates Results

Theorem

Let $x^\dagger \neq 0$ satisfy a Hölder source condition, g_α be regular, with finite qualification $\mu_0 < \infty$. $\mu^* = \min\{\mu, \mu_0\}$, $\tilde{\mu} = \min\{\mu, \mu_0 - \frac{1}{2}\}$.

- *Choose $\alpha^*_{y_\delta}$ by ψ_{QO} and let $y_\delta - y \in \mathcal{N}_1$*

$$\|x_{\alpha,\delta} - x^\dagger\| \leq C\delta^{\frac{2\mu^*}{2\mu^*+1}\min\{\frac{\mu}{\mu_0},1\}} \tag{20}$$

- *Choose $\alpha^*_{y_\delta}$ by ψ_{HR} let $y_\delta - y \in \mathcal{N}_2$*

$$\|x_{\alpha,\delta} - x^\dagger\| \leq C\delta^{\frac{2\mu^*}{2\mu^*+1}\min\{\frac{\mu}{\mu_0},1\}} \tag{21}$$

- *Choose $\alpha^*_{y_\delta}$ by ψ_{WR} let $y_\delta - y \in \mathcal{N}_2$ and $\mu_0 \geq \frac{1}{2}$, then*

$$\|x_{\alpha,\delta} - x^\dagger\| \leq C\delta^{\frac{2\tilde{\mu}}{2\tilde{\mu}+1}\min\{\frac{\mu}{\mu_0-\frac{1}{2}},1\}} \tag{22}$$

Oracle Type Estimates: QO

Oracle Estimate:

$$\|x_{\alpha^*_{y_\delta},\delta} - x^\dagger\| \leq C\left(\inf_\alpha \|x_\alpha - x_{\alpha,\delta}\| + \|x_\alpha - x^\dagger\|\right) \qquad (23)$$

Need additional condition on $x^\dagger$

Decay Conditions

Theorem

If $y - y_\delta \in \mathcal{N}_1$ and $x^\dagger$ satisfies

$$\int_0^t d\|E_\lambda x^\dagger\|^2 \leq C\int_t^\infty r_t^2(\lambda)d\|E_\lambda x^\dagger\|^2 \qquad (24)$$

then ψ_{QO} yields an oracle type estimate

Oracle Type Estimates; HR finite qualification

Oracle Estimate:

$$\|x_{\alpha^*_{y_\delta},\delta} - x^\dagger\| \leq C\left(\inf_\alpha \|x_\alpha - x_{\alpha,\delta}\| + \|x_\alpha - x^\dagger\|\right) \qquad (25)$$

Theorem

If $y - y_\delta \in \mathcal{N}_2$, $\mu_0 < \infty$,

$$\int_0^t d\|E_\lambda x^\dagger\|^2 \leq C\int_t^\infty r_t^2(\lambda) d\|E_\lambda x^\dagger\|^2 \qquad (26)$$

and

$$\int_0^t d\|F_\lambda(y - y_\delta)\|^2 \leq t\int_t^\infty \lambda^{-1} d\|F_\lambda(y - y_\delta)\|^2 \qquad (27)$$

then ψ_{HR} *yields an oracle type estimate*

Oracle Type Estimates; HR infinite qualification, WR

Oracle Estimate:

$$\|x_{\alpha^*_{y_\delta},\delta} - x^\dagger\| \leq C\left(\inf_{\alpha}\|x_\alpha - x_{\alpha,\delta}\| + \|x_\alpha - x^\dagger\|\right) \tag{28}$$

Theorem

If $y - y_\delta \in \mathcal{N}_2$, and additional to the previous Theorem let

$$\int_t^\infty \lambda d\|E_\lambda x^\dagger\|^2 \leq \frac{1}{t}\int_0^t d\|E_\lambda x d\|^2 \quad 0 \leq t \leq t_0 \tag{29}$$

then an oracle type estimate for ψ_{HR} for $\mu_0 = \infty$, and for ψ_{WR} for $\mu_0 \geq \frac{1}{2}$ holds.

Discussion: Noise Condition and Convergence

Noise Condition

- If singular values decay polynomially or exponentially and noise components $\sim$ polynomially then noise conditions are satisfied.
- Polynomial or exponentially ill-posed problems with random noise: noise condition are satisfied.
- **Not satisfied** for very weakly ill-posed problem: singular value decay logarithmically and noise components decay polynomially.

Interpretation: Noise condition means that *Noise should not be too smooth*.

Better:

Noise condition means *Noise should be sufficiently different from* $R(T)$.

If this holds than $\psi_{QO}, \psi_{HR}, \psi_{WR}$ yield convergence. ψ_{mL} cannot be treated within our analysis (nonconvergence ?)

Discussion: Convergence Rates

Under noise condition we obtain convergence rates under Hölder source conditions for $\psi_{QO}, \psi_{HR}, \psi_{WR}$.

Suboptimal rate with a loss of a factor $\frac{\mu}{\mu_0}$ for ψ_{QO}, ψ_{HR}, and $\frac{\mu}{\mu_0 - \frac{1}{2}}$ for ψ_{WR}

Optimal order for smoothness at saturation $\mu = \mu_0$ or $\mu = \mu_0 - \frac{1}{2}$.

Conclusion: Convergence rates result are only good for low saturating regularization schemes.

Discussion: Oracle Estimates

Much stronger estimates than rate result !
Rate estimates are worst case within smoothness class, but 'average' case can be much better

For saturating regularization schemes $\mu_0 < \infty$, ψ_{QR} and ψ_{HR} are about equivalent. ψ_{HR} needs an additional condition that the noise is not too bad (not too slowly decaying).

We need an condition on $x^\dagger$: Should not be too "rough". (Contrary to noise condition)
Not satisfied for logarithmic source conditions.

ψ_{WR} needs strong additional conditions !
For $\mu_0 = \infty$ ψ_{HR} need strong conditions as well.

Bad situations for heuristic parameter choice rules

Bad situations:

- Smooth noise (not random)
- rough exact solution
- Regularization method without qualification $\mu_0 = \infty$ (smoothness of $x^\dagger$ not close to qualification)

These bad situations should be taken into account when testing heuristic parameter choice rules

Comparison

Convergence: ψ_{HR}, ψ_{WR} slightly better than ψ_{QO}

Convergence rates (worst case in $x^{\dagger}$, Hölder source condition)

finite qualification, in general only suboptimal rates

Low smoothness: ψ_{WR}

Higher smoothness: ψ_{HR}, ψ_{QO}

Oracle estimates (optimal order, need restriction on $x^{\dagger}$)

finite qualification

ψ_{QO} slightly better than ψ_{HR}, both much better ψ_{WR}

infinite qualification

ψ_{QO} much better than ψ_{HR}, ψ_{WR}

Acknowledgment

Thanks for useful discussions to:
Andreas Neubauer, Shuai Lu, Peter Mathe, Dirk Lorenz

Recent References:

- S. Kindermann and A. Neubauer *On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization* Inv. Prob. Imag., 2, Nr. 2, pp 291-299, 2008
- A. Neubauer *The convergence of a new heuristic parameter selection criterion for general regularization methods* Inverse Probl. 24, No. 5, Article ID 055005 2008
- R. Palm. Numerical comparison of regularization algorithms for solving ill-posed problems (PhD thesis, supervised by U. Hämarik).
- S. Lu and P. Mathe, *Heuristic parameter selection based on functional minimization: Optimality and model function approach* WIAS preprint 1413
- B, Jin and D. Lorenz *Heuristic parameter-choice rules for convex variational regularization based on error estimates* [arXiv.org/abs/1001.5346]