Heuristic Parameter Choice Rules for Tikhonov Regularization with Weakly Bounded Noise

Stefan Kindermann and Kemal Raik
Industrial Mathematics Institute, Johannes Kepler University Linz, Linz, Austria

ABSTRACT
We study the choice of the regularization parameter for linear ill-posed problems in the presence of noise that is possibly unbounded but only finite in a weaker norm, and when the noise-level is unknown. For this task, we analyze several heuristic parameter choice rules, such as the quasi-optimality, heuristic discrepancy, and Hanke-Raus rules and adapt the latter two to the weakly bounded noise case. We prove convergence and convergence rates under certain noise conditions. Moreover, we analyze and provide conditions for the convergence of the parameter choice by the generalized cross-validation and predictive mean-square error rules.

ARTICLE HISTORY
Received 17 September 2018
Revised 2 April 2019
Accepted 2 April 2019

KEYWORDS
Generalized cross-validation; ill-posed problems; inverse problems; large noise; parameter choice rules; regularization

1. Introduction
Let $X$ and $Y$ be Hilbert spaces and $T : X \rightarrow Y$ be a compact linear operator. We consider the ill-posed problem

$$Tx = y,$$  

in which $T$ may have a nontrivial kernel and where we do not know $y$ exactly, but only noisy data $y^\delta = y + e$ are available. In contrast to the standard setting, the main focus of this article concerns the case of possibly unbounded noise, that is, $\delta := ||e||$ is possibly infinite. The latter may occur, for instance, in the case where we have white noise and $Y$ is the space of square summable sequences (cf. [1, 2]). It may be, however, that the noise is weakly bounded (cf. [3–10]), which we define as being whenever

$$\eta := ||(TT^*)^p(y^\delta - y)|| < \infty, \quad \text{for some} \quad p \in [0, \frac{1}{2}].$$  

The aforementioned references, besides [5], are restricted to the particular case in which $p = \frac{1}{2}$. Since $T$ is compact and $\dim \mathcal{R}(T) = \infty$, it follows that $\mathcal{R}(T)$ is non-closed, which implies that the generalized inverse
(see, e.g., [11]) \( T^\dagger \) is an unbounded operator. We therefore introduce regularization. We opt to employ Tikhonov regularization (cf. [12]) in which the regularized solution is given by

\[
x_{\delta} := (T^* T + \alpha I)^{-1} T^* y^\delta.
\]

We also denote \( x_\delta \) as the regularized solution with exact data. Note that by (2) and \( \eta \leq 1/2 \), \( x_\delta \) is well-defined. Furthermore, we shall assume henceforth that \( y \in D(T^\dagger) \). Then, in the case that \( y \) is non-attainable, that is, \( y \not\in \mathcal{R}(T) \), we may reduce to the attainable case by considering \( Tx = Qy \) where \( Q : Y \rightarrow \overline{\mathcal{R}(T)} \) is an orthogonal projection (cf. [13]).

Additionally, we denote \( \mathcal{Z} \) the Hilbert space to which the noise \( y^\delta - y \) belongs. It can be defined as the completion of \( \overline{\mathcal{R}(T)} \) with respect to \( \|\cdot\|_Y \) complemented by \( \mathcal{R}(T)^\perp \). Equivalently, in case of \( T^* \) injective, it could be defined as the negative Hilbert scale space \( \mathcal{X}_{-p} \) generated by the operator \( (T^*)^{-1} \); cf. [13, Section 8.4].

Our central aim is to approximate the best approximate solution \( x^\dagger = T^\dagger y \), such that \( x_\delta \) converges to \( x^\dagger \) in the weakly bounded noise case, that is, as \( \eta \rightarrow 0 \) for an appropriately selected \( \alpha \).

In the current setting, (cf. [4, 5, 9]) the balancing principle or modified discrepancy rules were suggested for the parameter choice. Note that these are a-posteriori rules which require knowledge of the noise level. In practical situations, such information is not normally available, and this motivates the need for so-called heuristic parameter choice rules in which the parameter is selected as the minimizer of a functional \( \psi : (0, ||T||^2) \times Y \rightarrow [0, \infty] \). In particular, if we define

\[
\tilde{\psi}(\alpha, y^\delta) := \begin{cases} 
\psi(\alpha, y^\delta), & \text{if } \alpha > 0, \\
\liminf_{\gamma \rightarrow 0} \psi(\gamma, y^\delta), & \text{if } \alpha = 0,
\end{cases}
\]

then

\[
\alpha_* := \arg\min_{\alpha \in [0, ||T||^2]} \tilde{\psi}(\alpha, y^\delta),
\]

which requires no knowledge of \( \eta \). Note that in case the functional exhibits multiple minimizers, we choose the largest one. The main objective of this article is the analysis of heuristic parameter choice rules in the weakly bounded noise (aka large noise) case.

The functionals \( \psi \) in this article may also be represented in terms of spectral theory:

\[
\psi^2(\alpha, y^\delta) = \int_0^{|T|^2} \Psi_\alpha(\lambda) \ d||F_\lambda y^\delta||^2,
\]

where \( \Psi_\alpha : (0, ||T||^2) \rightarrow \mathbb{R}_+ \) is a spectral filter function and \( \{F_\lambda\}_{\lambda} \) denotes the spectral family of \( TT^* \). For later reference, we also define \( \{E_\lambda\}_{\lambda} \) to be
the spectral family of $T^*T$. Note that in the following, $C$ will denote an arbitrary positive constant which need not be universally equal.

The article is organized as follows: in the proceeding section, we study and extend the classical heuristic parameter choice rules, namely, the quasi-optimality, heuristic discrepancy, and Hanke-Raus rules. We establish convergence rates under noise conditions similar to the strongly bounded noise case [14, 15]. In Section 3, we investigate known statistical rules in a deterministic framework, in particular, the generalized cross-validation rule. Since this is only defined in a discrete setting, we first analyze the theoretical (i.e. non-implementable) infinite-dimensional variant, that is, the predictive mean-square error functional.

2. Heuristic parameter choice rules

The standard method of approach to prove convergence rates for heuristic parameter choice rules is to estimate the data error from above by $\psi(\alpha, y^\delta)$ for which we also attain an estimate from above. One also estimates $\psi(\alpha, y)$ from above. If $\alpha_*$ is the minimizer of $\psi(\alpha, y^\delta)$, then

$$||x^\delta_2 - x^\dagger|| \leq ||x^\delta_2 - x_*|| + ||x_* - x^\dagger|| = O(\psi(\alpha_*, y - y^\delta) + ||x_* - x^\dagger||)$$

$$= O\left(\psi(\alpha, y^\delta) + \psi(\alpha_*, y) + ||x_* - x^\dagger||\right),$$

from which the derivation of the rates is quite standard.

Specifically, in this article, we consider heuristic rules based on the following $\psi$-functionals:

- The quasi-optimality functional (cf. [12])

$$\psi_{\text{QO}}(\alpha, y^\delta) := \alpha ||\frac{d}{d\alpha} x^\delta_2||; \quad (4)$$

- The modified heuristic discrepancy functional (cf. [16])

$$\psi_{\text{HD}}(\alpha, y^\delta) := \frac{1}{\alpha^{q+\frac{1}{2}}} ||(TT^*)^q \left(Tx^\delta_2 - y^\delta\right)||, \quad \text{where } q \geq p;$$

- The modified Hanke-Raus functional (cf. [16])

$$\psi_{\text{HR}}(\alpha, y^\delta) := \frac{1}{\alpha^{q+\frac{1}{2}}} \left((TT^*)^q \left(Tx^\delta_2 - y^\delta\right) + (TT^*)^q \left(Tx^\delta_2 - y^\delta\right)^2\right)^\frac{1}{2}, \quad \text{where } q \geq p.$$

where $x^\delta_2 := (T^*T + \alpha I)^{-1}(T^*y^\delta + \alpha x^\delta_2)$ is the second iterated Tikhonov solution.
In terms of (3), this corresponds to selecting
\[ \Psi_x(\lambda) := \frac{\lambda^{2q}x^m}{x^q(\lambda + x)^{m+1}} \ 	ext{with} \begin{cases} 
q \geq p \text{ and } m = 1, & \text{if } \psi = \psi_{\text{HD}}, \\
q \geq p \text{ and } m = 2, & \text{if } \psi = \psi_{\text{HR}}, \\
qu = \frac{1}{2} \text{ and } m = 3, & \text{if } \psi = \psi_{\text{QO}}. 
\end{cases} \]

Note that the heuristic discrepancy rule is sometimes confusingly also referred to as the Hanke-Raus rule (as the rules coincide for Landweber iteration). For clarity, it is preferable to name this method as the heuristic analog of the classical discrepancy rule. For advanced numerical implementations of these methods, see [17].

Note that our definitions of the heuristic discrepancy and Hanke-Raus functionals are generalizations of the usual ones. The usual functionals are obtained for the special case \( q = 0 \). The reason for this modification is that in the setting of weakly bounded noise, the discrepancy is possibly unbounded, and hence, the standard functionals need not be bounded either. Therefore, by introducing the operator \((TT^*)^q\), the functionals become finite if \( q \) is chosen larger than \( p \). This is a simple exercise to prove. Note that the quasi-optimality functional does not require any modification.

The drawback of heuristic parameter choice rules comes in the form of the so-called Bakushinskii veto, which states that choosing the parameter heuristically cannot lead to a convergent regularization method in the worst case (cf. [18]). Despite this, heuristic rules are still very often used with great success in practice. Motivated by this, it was shown that if one does not consider the worst case, heuristic rules may lead to convergent regularization methods. In particular, in [14, 15], additional noise conditions were postulated in order to estimate the data error as
\[ ||x_\delta^x - x_\lambda|| \leq C_\psi(x_\delta - y_\delta), \]
from which we can prove convergence of the method. As we will show in the subsequent sections (and as was proven for the bounded noise case in [14, 15]), the estimate (5) is obtained for the mentioned rules whenever we impose a noise condition \( y - y_\delta \in \mathcal{N}_\nu \), that is,
\[ \mathcal{N}_\nu := \left\{ e \in \mathcal{Z} : \nu + 1 \int_\alpha^{||x||^2} \lambda^{-1} d||F_\lambda e||^2 \leq C \int_0^\alpha \nu \lambda^\nu d||F_\lambda e||^2 \right\}, \]
where \( \nu = 1 \) for \( \psi = \psi_{\text{QO}} \) and \( \nu = 2q \) for \( \psi \in \{\psi_{\text{HD}}, \psi_{\text{HR}}\} \).

Let us state some simple examples, where a noise condition (6) holds, and, in particular, convince the reader that the assumption of weakly bounded noise is compatible with condition (6). Note that in the classical situation of (strongly) bounded noise, it has been verified that (6) is
satisfied in typical situations [15]. Moreover, for Colored Gaussian noise, (6) holds almost surely [19].

Suppose that $TT^*$ has eigenvalues $\{\lambda_i\}$ with polynomial decay, and we assume a certain polynomial decay or growth of the noise $e = y^\delta - y$ with respect to the eigenfunctions of $TT^*$, denoted by $\{u_i\}$:

$$\lambda_i = \frac{1}{i^\nu}, \quad \gamma > 0, \quad \text{and} \quad |\langle y^\delta - y, u_i \rangle|^2 = \frac{1}{i^{\beta}} \quad \beta \in \mathbb{R}, \tau > 0. \quad (7)$$

Then

$$||y^\delta - y||^2 = \tau \sum_{i=1}^{\infty} \frac{1}{i^{\beta}}, \quad \eta^2 = ||(TT^*)^p (y^\delta - y)||^2 = \tau \sum_{i=1}^{\infty} \frac{1}{i^{\beta+2p\gamma}}.$$  

If we consider the case of unbounded but weakly bounded noise, that is, $||y^\delta - y||^2 = \infty$ but $\eta < \infty$, the exponents $\beta, p$ should thus satisfy

$$\beta \leq 1 \quad \text{and} \quad \beta + 2p\gamma > 1, \quad \text{thus}, \quad \beta \in (1-2p\gamma, 1].$$

The inequality in (6) can then be written as

$$\tau x^{\nu+1} \sum_{1 \leq i \leq x^\frac{1}{\nu}} i^{-\beta} = \tau x^{\nu+1} \sum_{\lambda_i \geq x^\frac{1}{\nu}} \frac{1}{\lambda_i^{\beta}} |\langle (y^\delta - y), u_i \rangle|^2 \leq C \sum_{\lambda_i \geq x^\frac{1}{\nu}} \lambda_i^\nu |\langle (y^\delta - y), u_i \rangle|^2 = C \sum_{i \geq x^\frac{1}{\nu}} \frac{1}{i^{\nu+\beta}}.$$  

Defining $N_* = x^{-\frac{1}{\nu}}$, we have

$$\sum_{1 \leq i \leq x^{\frac{1}{\nu}}} i^{-\beta} \leq \int_1^{N_*} x^{\nu-\beta} \ dx \leq C \begin{cases} 1 & \text{if } \gamma - \beta > -1, \\ N_*^{-\beta+1} & \text{if } \gamma - \beta < -1, \end{cases}$$

and

$$\sum_{i \geq x^{\frac{1}{\nu}}} \frac{1}{i^{\nu+\beta}} \sim \int_{N_*}^{\infty} \frac{1}{x^{\nu+\beta}} \ dx \sim \frac{C}{N_*^{\nu+\beta-1}} \begin{cases} 1 & \text{if } \gamma \nu + \beta > 1, \\ \infty & \text{if } \gamma \nu + \beta \leq 1. \end{cases}$$

Since $\chi = N_*^{-\gamma}$, we arrive at the sufficient inequality

$$N_*^{-\gamma(\nu+1)+\gamma-\beta} \leq C N_*^{-\gamma-\beta},$$

in the case that $\gamma - \beta > -1$ and $\gamma \nu + \beta > 1$. Since the exponents match, the noise condition is then satisfied. If $\gamma \nu + \beta \leq 1$, then the inequality is clearly satisfied because of the divergent right-hand side. Thus, the noise condition holds for

$$\beta < \gamma + 1.$$  

Roughly speaking, this means that the noise should not be too regular (relative to the smoothing of the operator). In particular, the deterministic model of white noise, where $\beta = 0$ (no decay) satisfies a noise condition if
the operator is smoothing. Most importantly, the assumption of a noise condition (6) is compatible with a weakly bounded noise situation.

In the latter sections, we also consider the predictive mean-square error (PMS) functional \([20, 21]\)

\[
\psi_{\text{PMS}}(x, y^\delta) := ||Tx^\delta - y||.
\]

This is not an implementable parameter choice rule per se as it involves the (unknown) exact data \(y\). The reason for opting to study this functional is its relation to the generalized cross-validation functional, which is one of our main aims. For ill-conditioned problems \(T_n x = y_n\), where \(T_n : X \to \mathbb{R}^n\), the generalized cross-validation (GCV) functional \([21]\) is given by

\[
\psi_{\text{GCV}}(x, y_n^\delta) := \frac{1}{\rho(x)} ||T_n x_n^\delta - y_n^\delta||,
\]

with \(\rho(x) := \frac{2}{n} \text{tr}\{(T_n T_n^* + \alpha I)^{-1}\}\). Its relation to \(\psi_{\text{PMS}}\) is that for i.i.d. noise, the expected value of \(\psi_{\text{GCV}}(x, y_n^\delta) - ||e||^2\) is an estimator for the expected value of the predictive mean-square error functional squared, as has been shown by Wahba \([21]\). For the numerical treatment of the GCV method, see, for example \([22]\).

### 2.1. Convergence analysis

The convergence analysis of regularization methods with standard (non-heuristic) parameter choice rules in the weakly bounded noise setting is well established: for instance, in the present setting, one can easily show, as in \([4]\), that

\[
||x^\delta_\alpha - x^\dagger|| \to 0,
\]

if one chooses \(\alpha_\star\) such that \(\alpha_\star \to 0\) and \(\eta^{2+p}/\alpha_\star \to 0\) as \(\eta \to 0\). Therefore, even in the presence of large noise, one may obtain a convergent regularization method.

We are also interested in deriving rates of convergence. To this end, we assume throughout that the best approximate solution \(x^\dagger \in X\) satisfies the source condition:

\[
x^\dagger \in \mathcal{R}((T^* T)^{\mu}) \iff x^\dagger = (T^* T)^{\mu} \omega, \quad ||\omega|| < \infty, \quad 0 \leq \mu \leq 1,
\]

which one can think of as a kind of smoothness condition on the solution.

The following error estimates are courtesy of \([5]\) (cf. also \([4, 9]\)):

**Proposition 1.** Let \(x^\dagger\) satisfy (8). Then

\[
||x^\delta_\alpha - x_\star|| \leq C \frac{\eta}{\alpha^{p+\frac{1}{2}}}, \quad ||x_\star - x^\dagger|| \leq C \alpha^\mu, \quad \mu \leq 1,
\]

(9)
\[
\|T(x_\delta^\lambda - x_\lambda)\| \leq C \frac{\eta}{\lambda^6}, \quad \|Tx_\lambda - y\| \leq C\lambda^{1/2}, \quad \mu \leq \frac{1}{2},
\]

(10)

for all \( \lambda \in (0, \|T\|^2) \).

This proposition also illustrates the fact that convergence rates for Tikhonov regularization do not improve for \( \mu \geq 1 \), which is the well-known saturation effect (cf. [13]). This is also the reason why we do not assume a source condition in (8) with \( \mu > 1 \).

We now consider an a-priori parameter choice yielding a so-called optimal (order) rate. Thereafter, we will utilize this a-priori parameter choice strategy to deduce convergence rates with respect to the heuristic parameter choice rules. In particular, if \( x^\dagger \) satisfies the source condition (8), then using the estimates of the previous proposition, one can estimate the total error as

\[
\|x_\delta^\lambda - x^\dagger\| \leq C\lambda^{\mu} + C\frac{\eta}{\lambda^{p+\frac{1}{2}}} = \mathcal{O}\left(\eta^{\frac{2p}{2p+2p+1}}\right),
\]

(11)

which follows by taking the infimum over all \( \lambda \). In particular, one obtains that

\[
\lambda_{\text{opt}} \approx \eta^{\frac{2}{2p+2p+1}},
\]

is the so-called optimal (order) parameter choice.

For the following analysis, we state a standard estimate for spectral filter functions: for \( t \geq 0 \), there is a constant \( C \) such that for all nonnegative \( \lambda, \lambda \)

\[
\frac{\lambda^t}{(\lambda + \lambda)^{t}} \leq C\left\{ \begin{array}{ll}
C & \text{if } t \leq 1 \\
C^{-1} & \text{if } t \geq 1
\end{array} \right. = \frac{C}{\lambda_{\max}(1-t,0)},
\]

(12)

Lemma 1. Assume that \( x^\dagger \) satisfies the source condition (8).

- If \( y - y^\delta \in N_1 \), then there exists a positive constant \( C \) such that
  \[
  C\|x_\delta^\lambda - x_\lambda\| \leq \psi_{\text{QO}}(\lambda, y - y^\delta) \leq \|x_\delta^\lambda - x_\lambda\|,
  \]
  (13)
  \[
  \psi_{\text{QO}}(\lambda, y) \leq \|x_\lambda - x^\dagger\|;
  \]
  (14)
- If \( Q(y - y^\delta) \in N_{2q} \), \( q \leq \min\{p + 1, \frac{1}{2} - \mu\} \). Then there exists positive constants such that
  \[
  \|x_\delta^\lambda - x_\lambda\| \leq \psi_{\text{HD}}(\lambda, y - y^\delta) \leq C\frac{\eta}{\lambda^{p+\frac{1}{2}}},
  \]
  (15)
  \[
  \psi_{\text{HD}}(\lambda, y) \leq C\lambda^{\mu};
  \]
  (16)
- Let \( Q(y - y^\delta) \in N_{2q}, q \leq \min\{p + \frac{3}{2}, 1 - \mu\} \). Then there exists positive constants such that
\[ C||x_2^\delta - x_\lambda|| \leq \psi_{HR}(x, y-y^\delta) \leq C \frac{\eta}{q^{p+\frac{1}{2}}}, \quad (17) \]
\[ \psi_{HR}(x, y) \leq Cx^\mu, \quad (18) \]

for all \( x \in (0, ||T||^2) \).

**Proof.** The estimates (13, 14) do not require the weakly boundedness condition on the noise and therefore the proofs are nigh on identical to the ones found in [15]. For \( m > 0 \) and with (12) we obtain
\[
\frac{\lambda^{2q}}{\lambda^{2q}(\lambda + \alpha)^{(m+1)}} \leq \frac{\lambda^{2p}}{\lambda^{2q-m}\left(\frac{2(q-m)}{\lambda + \alpha}\right)^{(m+1)}} \]
\[
\leq \frac{1}{\lambda^{2q-m+\max\left\{1-\frac{2q-m}{m+1},0\right\}(m+1)}} = C\frac{\lambda^{2p}}{\lambda^{2q-m+\max\left\{1-2p,2q-m\right\}}} = C\frac{\lambda^{2p}}{\lambda^{1+2p}},
\]
if \( q < \frac{m+1}{2} + p \). Moreover,
\[
\frac{\lambda^{2q}}{\lambda^{2q}(\lambda + \alpha)^{(m+1)}} \lambda^{1+2\mu} \leq \frac{1}{\lambda^{2q-m+\max\left\{1-\frac{2q-m}{m+1},0\right\}(m+1)}} = C\frac{\lambda^{2p}}{\lambda^{2q-m+\max\left\{2\mu,2q-m\right\}}} \leq C\lambda^{2\mu},
\]
if \( q < \frac{m}{2} - \mu \). Using the spectral representation, the upper estimates (15) (using \( m = 2 \)) and (17) (using \( m = 3 \)) follow from the first estimate, while (16) and (18) are obtained similarly by the second one.

For the lower bound, we estimate
\[
\psi^2_{HD}(x, y-y^\delta) = \frac{1}{\alpha^{2q+1}}\int_0^{||T||^2} \lambda^{2q} \frac{x^2}{(\lambda + \alpha)^2} \, d||F_\lambda Q(y-y^\delta)||^2
\]
\[
\geq C\frac{1}{\alpha^{2q+1}}\int_0^x \lambda^{2q} \, d||F_\lambda Q(y-y^\delta)||^2 + C\frac{1}{\alpha^{2q-1}}\int_x^{||T||^2} \lambda^{2q-2} \, d||F_\lambda Q(y-y^\delta)||^2,
\]
(19)

for all \( x \in (0, ||T||^2) \).

Conversely,
\[
||x_2^\delta - x_\lambda||^2 = \int_0^x \frac{\lambda}{(\lambda + \alpha)^2} \, d||F_\lambda Q(y-y^\delta)||^2
\]
\[
\leq C\frac{1}{\alpha} \int_0^x \lambda \, d||F_\lambda Q(y-y^\delta)||^2 + C\int_x^{||T||^2} \lambda^{-1} \, d||F_\lambda Q(y-y^\delta)||^2.
\]
(20)

Since \( 2q-1 \leq 0 \), we observe that the term with \( \int_0^x \) in the above inequality is bounded by the corresponding term in (19). Thus, using the noise condition, the second term can be bounded by the first one of (19).
For the lower bound of the modified Hanke-Raus functional, we can estimate
\[
\frac{\lambda^{2q}}{\alpha^{2q} (\lambda + \alpha)^3} \geq C \begin{cases} 
\frac{\lambda^{2q}}{\alpha^{2q+1}} & \text{if } \lambda \leq \alpha, \\
\frac{\lambda^{2q-3}}{\alpha^{2q-2}} & \text{if } \lambda \geq \alpha.
\end{cases}
\]

Now, using \(N_{2q}\) and (20), we can estimate \(||x^\beta - x^\alpha||\) by the part of \(\psi_{HR}(x^\alpha, y-y^\beta)\) restricted to \(\lambda \leq \alpha\). The part for \(\lambda \geq \alpha\) can then be estimated from below by 0. \(\square\)

Note that our parameters satisfy \(\mu \geq 0\) and \(p \in [0, \frac{1}{2}]\), hence, the restrictions on the parameter \(q\) reduce to \(q \leq \frac{1}{2} - \mu\) or \(q \leq 1 - \mu\) for the heuristic discrepancy or the Hanke-Raus rules, respectively. Since by definition \(q \geq p\) must hold in any case, we have as a restriction to the smoothness index that \(\mu \in [0, \frac{1}{2} - p]\) for \(\psi_{HD}\) and \(\mu \in [0, 1 - p]\) for \(\psi_{HR}\), respectively. Only then, there exists a possible choice for \(q\) that satisfies the conditions of the previous lemma. Observe that the interval for \(\mu\) is smaller for \(\psi_{HD}\), which also illustrates a saturation effect of the discrepancy-based rules which is well-known in the standard noise case, that is, \(p = 0\).

**Theorem 1.** Assume that \(x^\dagger\) satisfies the source condition (8) and additionally that
\[
\begin{aligned}
&y-y^\beta \in N_1 \text{ and } T^*y \neq 0 & &\text{if } \psi = \psi_{QO}, \\
&Q(y-y^\beta) \in N_{2q}, \mu \in [0, \frac{1}{2} - p], q \in [p, \frac{1}{2} - \mu] \text{ and } (TT^*)^qQy \neq 0 & &\text{if } \psi = \psi_{HD}, \\
&Q(y-y^\beta) \in N_{2q}, \mu \in [0, 1 - p], q \in [p, 1 - p] \text{ and } (TT^*)^qQy \neq 0 & &\text{if } \psi = \psi_{HR}.
\end{aligned}
\]

Then
\[
||x^\beta_{x^\alpha} - x^\dagger|| = \begin{cases} 
O(\eta^{\frac{2\mu}{2\mu + 2p + 1}}), & \text{if } \psi = \psi_{QO}, \\
O(\eta^{\frac{2\mu}{2\mu + 2p + 1}}(\frac{2\mu}{1 - 2q})), & \text{if } \psi = \psi_{HD}, \\
O(\eta^{\frac{2\mu}{2\mu + 2p + 1}}(\frac{2\mu}{1 - \eta})), & \text{if } \psi = \psi_{HR},
\end{cases}
\]
for \(\eta\) sufficiently small.

**Proof.** As in the proof of the previous lemma, we treat the different parameter choice rules separately:

- From the definition of \(x^\alpha\) and the triangle inequality, it follows, with \(\alpha = \eta^{\frac{2\mu}{2\mu + 2p + 1}}\), that
\[ \psi_{QO}(\alpha, y^\delta) \leq \psi_{QO}^2(\alpha, y^\delta) \leq \left( \psi_{QO}(\alpha, y^\delta - y) + \psi_{QO}(\alpha, y) \right)^2 \]
\[ \leq 2||x_\alpha - x^\dagger||^2 + 2||x_\alpha^\delta - x_\alpha||^2 \leq Cx^2\eta + C\eta^2 = O\left( \left[ \eta^{\frac{2\mu}{2\mu+2p+1}} \right]^2 \right). \]

By the triangle inequality, (13) and (14) of Lemma 1,
\[ ||x_\alpha^\delta - x^\dagger|| \leq ||x_\alpha - x^\dagger|| + ||x_\alpha - x_\alpha^\delta|| = O(\left(||x_\alpha - x^\dagger|| + \psi_{QO}(\alpha, y - y^\delta)\right)) \]
\[ \leq O\left(||x_\alpha - x^\dagger|| + \psi_{QO}(\alpha, y^\delta) + \psi_{QO}(\alpha, y)\right) = O\left(\alpha^\mu + \left[ \eta^{\frac{2\mu}{2\mu+2p+1}} \right]^2 \right). \]

Note that
\[ \psi_{QO}(\alpha, y^\delta) \geq x^2 \int_0^{||T||^2} \frac{\lambda}{(\lambda + ||T||^2)^4} \, d||F_2y^\delta||^2 \geq \alpha^2 \frac{1}{(2||T||^2)^4} \int_0^{||T||^2} \lambda \, d||F_2y^\delta||^2 \]
\[ \geq \alpha^2 \frac{1}{(2||T||^2)^4} \left(||T^*y|| - ||TT^*||^{\frac{1}{2-p}}\eta \right)^2 \geq Cx^2, \]
\[ \text{(21)} \]
for all \( \alpha \in (0, ||T||^2) \) and \( \eta \) sufficiently small. Hence for \( \alpha = \alpha_* \), it follows that \( \alpha_* \leq C\eta^{\frac{2\mu}{2\mu+2p+1}} \). Therefore, we may deduce that
\[ ||x_\alpha^\delta - x^\dagger|| = O\left(\eta^{\frac{2\mu}{2\mu+2p+1}}\right), \]
for \( \eta \) sufficiently small.

- Note that from \((TT^*)^qQy \neq 0\), we may conclude, as in (21), that
\[ \alpha_* \leq C\left(\psi_{HD}(\alpha, y^\delta)^{\frac{1}{1-q}}\right) = O\left(\eta^{\left(\frac{2\mu}{2\mu+2p+1}\right)^{\frac{1}{1-2q}}}\right). \]

Then it follows, as above, from (15) and (16), that
\[ ||x_\alpha^\delta - x^\dagger|| \leq ||x_\alpha - x^\dagger|| + ||x_\alpha - x_\alpha^\delta|| = O\left(\alpha^\mu + \psi(\alpha, y - y^\delta)\right) \]
\[ = O\left(\alpha_*^\mu + \alpha^\mu + \eta^{\frac{1}{\alpha^2 + p}}\right) = O\left(\eta^{\left(\frac{2\mu}{2\mu+2p+1}\right)^{\frac{1}{1-2q}}} + \eta^{\frac{2\mu}{2\mu+2p+1}}\right), \]
for \( \eta \) sufficiently small.

- One may similarly verify that if \(||(TT^*)^qQy|| \geq C\), then
\[ \alpha_* \leq C\psi_{HR}(\alpha, y^\delta)^{\frac{1}{1-q}}. \]

Therefore,
For the quasi-optimality rule, one may notice that the above convergence rates are optimal for the saturation case $\mu = 1$, but they are only suboptimal for $\mu < 1$ (similarly as in [14]).

Let us further discuss the assumptions in this theorem: for the modified heuristic discrepancy rule, the first condition on $q$ is not particularly restrictive. However, the requirement $q \leq \frac{1}{2} - \mu$ implies that $\mu \leq \frac{1}{2} - q$, which means that we obtain a saturation at $\mu = \frac{1}{2} - q$. This is akin to the bounded noise case ($q = 0$), where this method saturates at $\mu = \frac{1}{2}$. It is well known that a similar phenomenon occurs for the non-heuristic analog of this method, namely the discrepancy principle.

In contrast to the modified discrepancy rule, we observe that the saturation for the modified Hanke-Raus rule occurs at $\mu = 1 - q$. Hence, again analogous to the bounded noise case (and to the non-heuristic case), the modified Hanke-Raus method yields convergence rates for a wider range of smoothness classes.

We may, however, impose an additional condition in order to achieve an optimal convergence rate. More specifically, we impose the following regularity condition on the best approximate solution, $x^\dagger \in X$:

$$\alpha^2 \int_{x}^{\infty} \lambda^{-2} \, d\|E_j x^\dagger\|^2 \geq C \int_{0}^{x} d\|E_j x^\dagger\|^2. \quad (22)$$

This condition was also used in [14, 15] where it was shown that it is often satisfied.

**Theorem 2.** Let the assumptions of the previous theorem hold and let $\alpha$, be selected according to either the quasi-optimality, modified heuristic discrepancy, or the modified Hanke-Raus rule. Then, assuming the regularity condition (22), it follows that

$$||x^\delta_{x^\dagger} - x^\dagger|| = O(\eta^\frac{2\mu}{n + 2p + 1}),$$

for $\eta$ sufficiently small.

**Proof.** The proof for the quasi-optimality rule is identical to the one found in [14, 15]; therefore, we omit it. For the modified heuristic discrepancy rule and the Hanke-Raus rule, we show that the regularity condition implies that $\psi(x, y) \geq C ||x_{x^\dagger} - x^\dagger||$. Recall that
\[
\|x_\lambda - x^\dagger\|^2 = \int_0^{||T||^2} \frac{\lambda^2}{(\lambda + \lambda^2)^2} \, d\|E_i x^\dagger\|^2 \leq C \int_0^{||T||^2} \frac{1}{\lambda^2} \, d\|E_i x^\dagger\|^2.
\]

For \( \lambda \geq \alpha \) we have the following estimate for \( m > 0 \) with a constant \( C_{q,m} \):
\[
\left( \frac{\lambda}{\alpha} \right)^{2q} \frac{\lambda \alpha^m}{(\lambda + \alpha)^{m+1}} \geq C_{q,m} \lambda^m \alpha^m \geq C_{q,m} \lambda^{m+1} \frac{\alpha^{m+1}}{(\lambda + \alpha)^{m+1}}.
\]

Thus, for \( \psi = \psi_{HD} \) taking \( m = 1 \) and for \( \psi = \psi_{HR} \) taking \( m = 2 \), we obtain
\[
\psi^2(\alpha, y) \geq \int_0^{||T||^2} \left( \frac{\lambda}{\alpha} \right)^{2q} \frac{\lambda \alpha^m}{(\lambda + \alpha)^{m+1}} \, d\|E_i x^\dagger\|^2 \geq \int_0^{||T||^2} \left( \frac{\lambda}{\alpha} \right)^{2q} \frac{\lambda \alpha^m}{(\lambda + \alpha)^{m+1}} \, d\|E_i x^\dagger\|^2.
\]

By the regularity condition, the first integral in the upper bound in (23) can be estimated by the second part which agrees up to a constant with the lower bound for \( \psi(\alpha, y) \) in both cases. In the proof of Theorem 1, the estimate \( \|x_\alpha - x^\dagger\| \leq C\alpha^m \) can then be replaced by \( \|x_\alpha - x^\dagger\| \leq C\psi(\alpha, y) \), which leads to the optimal rate.

3. PMS and GCV

In this section we study the generalized cross-validation and its infinite-dimensional analog, the predictive mean-square error (PMS), in a deterministic framework. Note that the PMS is not an implementable heuristic rule as it requires the non-available exact data. Rather, it is considered here as—an only theoretically available—limit of the GCV functional in case of infinite dimensions, and it is analyzed in order to obtain an initial understanding of the GCV rule and to motivate the regularity conditions and noise conditions used below for the latter rule. In the following, we will state a noise and regularity condition from which we derive convergence rates in case \( \alpha \) is chosen as the minimizer of the predictive mean-square error functional.

3.1. The predictive mean-square error

The predictive mean-square error functional differs from the previous ones in the sense that it has different upper bounds. In fact, from (10), one immediately finds that
\[ \psi_{\text{PMS}}^2(\alpha, y^\delta) \leq C \frac{\eta^2}{2^p} + C x^{2\mu+1}, \]

for \( \mu \leq \frac{1}{2} \). The minimum of the upper bound is again obtained for \( \alpha = \alpha_{\text{opt}} = \mathcal{O}(\eta^{2/2\mu+1}) \), but the resulting rate is of the order

\[ \psi_{\text{PMS}}^2(\alpha, y^\delta) \leq C \left[ \frac{1}{2^{2\mu+1}} \right]^2, \]

which agrees with the optimal rate for the error in the T-norm, \(|x_\alpha^\delta - x^\gamma^\delta| \|_T := \| T(x_\alpha^\delta - x^\gamma^\delta) \|_T \). Thus, for this method, it is not reasonable to bound the functional \( \psi_{\text{PMS}} \) by expressions involving \( \| x_\alpha^\delta - x^\gamma^\delta \| \) or \( \| x_\alpha^\delta - x^\gamma^\delta \| \). Rather, we try to directly relate the selected regularization parameter \( \alpha \) to the optimal choice \( \alpha_{\text{opt}} \):

To do so, we need some estimates from below, although in this case, we will need to introduce a noise condition of a different type and an additional condition on the exact solution.

**Lemma 2.** Suppose that there exists a positive constant \( C \) such that \( y^\delta - y \in Z \) satisfies

\[ \int_{\mathbb{R}} |T|^2 \, d |F_0 Q(y - y^\delta)|^2 \geq C \frac{\eta^2}{2^{2p-\varepsilon}}, \quad (24) \]

for all \( \alpha \in (0, \|T\|^2) \) and \( \varepsilon > 0 \) small. Then

\[ \|T(x_\alpha^\delta - x^\gamma^\delta)\| \geq C \frac{\eta}{2^{p-\frac{\varepsilon}{2}}}. \]

**Proof.** From (24), one can estimate

\[ \|T(x_\alpha^\delta - x^\gamma^\delta)\|^2 = \int_0^{\|T\|^2} \frac{x^2}{(x + \lambda)^2} \, d |F_0 Q(y - y^\delta)|^2 \geq \int_0^{\|T\|^2} |T|^2 \, d |F_0 Q(y - y^\delta)|^2 \geq C \frac{\eta^2}{2^{2p-\varepsilon}}. \]

Let us exemplify condition (24): for the case in (7), we have that

\[ \int_{\mathbb{R}} d |F_0 Q(y - y^\delta)|^2 = \sum_{1 \leq i \leq N_\varepsilon} \frac{1}{i^\beta} \sim \int_1^{N_\varepsilon} \frac{1}{x^\beta} \, dx = \left\{ \begin{array}{ll} CN_\varepsilon^{1-\beta} & \text{if } 1-\beta > 0, \\ C & \text{if } 1-\beta < 0, \end{array} \right. \]

with \( N_\varepsilon = \frac{1}{d^\varepsilon} \). This gives that the left-hand side is of the order of \( \alpha^{-\frac{1-\beta}{\gamma}} \). For (24) to hold true, we require that \( \frac{1-\beta}{\gamma} \geq 2p - \varepsilon \), which means that

\[ 1 + \varepsilon \gamma \geq \beta + 2p \gamma. \]

If we now choose \( p \) close to the smallest admissible exponent for the weakly bounded noise condition, that is, \( 2p \gamma = 1-\beta + \varepsilon \gamma \), with \( \varepsilon \) small, then the condition holds. In other words, our interpretation of the stated
noise condition means that $\|((TT^*)^p(y^d - y))\| < \infty$ and $p$ is selected as the minimal exponent such that this holds. This noise condition automatically excludes the (strongly) bounded noise case. It can easily be seen that for strongly bounded noise $\|y^d - y\| < \infty$, the method fails as it selects $\alpha_* = 0$. The example also shows that the desired inequality with $\varepsilon = 0$ cannot be achieved.

**Theorem 3.** Let $\mu \leq \frac{1}{2}$, $\alpha_*$ be the minimizer of $\psi_{\text{PMS}}(x, y^d)$, assume that the noise satisfies (24) and that $Tx^t \neq 0$. Then

$$
||x^d_{x_*} - x^\dagger|| \leq \begin{cases} 
C\eta \left( \frac{2p + 2p + 1}{2p + 1} \right)^{\frac{2p + 1}{2}}, & \text{if } \alpha_* \geq \alpha_{\text{opt}}, \\
C\eta \left( \frac{2p + 2p + 1}{2p + 1} \right)^{-\frac{2p + 1}{2}}, & \text{if } \alpha_* \leq \alpha_{\text{opt}}.
\end{cases}
$$

If additionally for some $\varepsilon > 0$,

$$
\int_x ||T||^2 \lambda^{2\mu - 1} \, d||E_{\lambda}\omega||^2 \geq C\varepsilon^{2\mu - 1 + \varepsilon_2},
$$

then for the first case we have

$$
||x^d_{x_*} - x^\dagger|| \leq C\eta \left[ \frac{2p + 2p + 1}{2p + 1} \right]^{\frac{2p + 1}{2}}, \quad \text{if } \alpha_* \geq \alpha_{\text{opt}},
$$

**Proof.** If $\alpha_* \geq \alpha_{\text{opt}}$, it follows from $Tx^t \neq 0$ that

$$
||Tx^t - y||^2 \geq C\alpha^2,
$$

and if (25) holds, then one even has that

$$
||T(x^t_{x_*} - x^t)||^2 \geq \int_x ||T||^2 \lambda^{1 + 2\mu} \frac{\alpha^2}{(\alpha + \lambda)^2} \, d||E_{\lambda}\omega||^2 \geq \alpha^2 \int_x \lambda^{2\mu - 1} \, d||E_{\lambda}\omega||^2 \geq C\alpha^{2\mu + 1 + \varepsilon_2}.
$$

Since $\alpha \mapsto ||T(x^d_{x_*} - x^t)||^2$ is a monotonically decreasing function and using Young’s inequality, we may obtain that

$$
C\alpha_{\text{opt}}^t \leq ||T\left(x^d_{x_{\text{opt}}} - x^t_{x_{\text{opt}}}\right)||^2 + ||Tx^t_{x_{\text{opt}}} - y||^2 \leq C\eta^{\frac{2p + 1}{2p + 1}}^{\frac{2p + 1}{2}},
$$

that is,

$$
\alpha_* \leq C\eta^{\left( \frac{2p + 1}{2p + 1} \right)^{\frac{2p + 1}{2}}},
$$

where $t = 2$ or $t = 2\mu + 1 + \varepsilon_2$ if (25) holds.

If $\alpha_* \leq \alpha_{\text{opt}}$, then we may bound the functional from below as

$$
\psi^2_{\text{PMS}}(x, y^d) \geq \frac{1}{2} ||T\left(x^d_{x_*} - x^t\right)||^2 - ||Tx^t - y||^2,
$$

for all $x \in (0, ||T||^2)$, which allows us to obtain
that is, by Lemma 2,
\[ C \frac{\eta^2}{\alpha_{n}^{2p-\varepsilon}} - C \alpha_{\ast}^{2\mu+1} \leq \frac{1}{2} \| T(x_{\ast}^{\delta} - x_{\ast}) \|^2 - \| Tx_{\ast} - y \|^2 \leq C \left[ \eta^{2\mu+1} \right]^2. \]

Now, from \( \alpha_{\ast} \leq \alpha_{\text{opt}} \), we get
\[ C \frac{\eta^2}{\alpha_{n}^{2p-\varepsilon}} \leq C \left[ \eta^{2\mu+1} \right]^2 + C \alpha_{\ast}^{2\mu+1} \leq C \left[ \eta^{2\mu+1} \right]^2 + C \alpha_{\text{opt}}^{2\mu+1} \leq C \left[ \eta^{2\mu+1} \right]^2, \]
that is,
\[ \alpha_{n}^{2p-\varepsilon} \geq C \left[ \eta^{2\mu+1} \right]^2 \iff \alpha_{\ast} \geq C \eta^{2\mu+1} 2p \frac{2p}{1}. \]

Then inserting the respective bounds for \( \alpha_{\ast} \) into (11) yields the desired rates.

Condition (25) can again be verified as we did for the noise condition for some canonical examples. The inequality with \( \epsilon_2 = 0 \) does not usually hold. The condition can be interpreted as the claim that \( x^{\dagger} \) satisfies a source condition with a certain \( \mu \) but this exponent cannot be increased, that is, \( x^{\dagger} \notin \mathcal{R}((T^*T)^{\mu+\epsilon}) \). A similar condition was used by Lukas in his analysis of the generalized cross-validation rule [20].

The theorem shows that we may obtain almost optimal convergence results but only under rather restrictive conditions. Moreover, the method shows a saturation effect at \( \mu = \frac{1}{2} \) comparable to the discrepancy principles.

### 3.2. The generalized cross-validation rule

The generalized cross-validation rule was proposed and studied in particular by Wahba [21], and it is most popular in a statistical context but less so for deterministic inverse problems. It is derived from the cross-validation method by combining them with certain weights. Importantly, it was shown in [21] that the expected value of the generalized cross-validation functional converges to the expected value of the PMS-functional as the dimension tends to infinity. This is why, in the last section, we studied \( \psi_{\text{PMS}} \) in detail.

#### Proposition 2. Let \( \sup_n \text{tr}(T_n^*T_n) < \infty \); then it follows that the weight \( \rho(\alpha) \) in \( \psi_{\text{GCV}} \) is monotonically increasing with \( \rho(0) = 0 \) and bounded with \( \rho(\alpha) \leq 1 \). Furthermore, for \( \alpha > 0 \), it follows that \( \rho(\alpha) \to 1 \) as the dimension \( n \to \infty \).
Proof. Observe that

\[
\rho(\alpha) = \frac{1}{n} \text{tr} \left( (T_n^* T_n + \alpha I)^{-1} \right) = \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_i}{\alpha + \lambda_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\alpha + \lambda_i} = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_i}{\alpha + \lambda_i}.
\]

Moreover, it is clear that \( \rho(\alpha) \leq 1 \) for all \( \alpha > 0 \), since the second term is positive. Additionally, since the second term converges to 0 as \( n \to \infty \), it also follows that \( \rho(\alpha) \to 1 \) as \( n \to \infty \). The fact that \( \rho(0) = 0 \) is obvious. □

This is also the reason why one has to study the GCV in terms of weakly bounded noise. The limit \( \lim_{n \to \infty} \psi_{\text{GCV}} \) tends pointwise to the residual \( ||Tx^\d_t - y^\d|| \), which in the bounded noise case does not yield a reasonable parameter choice as then \( \alpha_\ast = 0 \) is always chosen.

Note that in a stochastic context, and using the expected value of \( \psi_{\text{GCV}} \), a convergence analysis has been done by Lukas [20]. In contrast, we analyze the deterministic case.

We now consider the ill-conditioned problem

\[
T_n x = y_n,
\]

where we only have noisy data \( y_n^\d \in \mathbb{R}^n \).

We impose a discretization independent source condition, that is,

\[
x^\d = (T_n^* T_n)^\mu \omega, \quad ||\omega|| \leq C, \quad 0 < \mu \leq 1,
\]

where \( C \) does not depend on the dimension \( n \). Furthermore, let us restate some definitions for this discrete setting:

\[
\delta_n := ||y_n^\d - y_n||, \quad \eta^2 := \sum_{i=1}^{n} \lambda_i^{2p} |\langle y_n^\d - y_n, u_i \rangle|^2.
\]

Note that in an asymptotically weakly bounded noise case, we might assume that \( \eta \) is bounded independent of \( n \) while \( \delta_n \) might be unbounded as \( n \) tends to infinity.

Moreover, we impose a noise condition of similar type as for the predictive mean-square error

\[
\sum_{\lambda_i \geq \alpha} |\langle y_n^\d - y_n, u_i \rangle|^2 \geq C \sum_{\lambda_i \geq \alpha} \eta^2, \quad \text{for all } \alpha \in I,
\]

where \( C \) does not depend on \( n \). Note that in the discrete case, one must restrict the noise condition to an interval with \( I = [\alpha_{\text{min}}, ||T||^2] \) with \( \alpha_{\text{min}} > 0 \).
Similarly, we state a regularity condition
\[
\sum_{\lambda_i \geq \alpha} \lambda_i^{2\mu - 1} |\langle \omega, v_i \rangle|^2 \geq C\alpha^{2\mu - 1 + \varepsilon_2} \quad \text{for all } \alpha \in I,
\]
(28)

where \( \{v_i\} \) denote the eigenfunctions of \( T^*T \).

In order to deduce convergence rates, we look to bound the functional from above as we did for the other functionals in the previous sections:

**Lemma 3.** For \( y^\delta_n \in \mathbb{R}^n \), there exist positive constants such that
\[
\psi_{GCV}(\alpha, y_n) \leq \frac{C}{\rho(\alpha)} C\alpha^{2\mu + 1}, \quad \mu \leq \frac{1}{2},
\]
(29)
\[
\psi_{GCV}(\alpha, y^\delta_n - y_n) \leq \frac{C}{\rho(\alpha)} \delta_n^2, \quad \mu \leq \frac{1}{2}, \quad \text{hence,}
\]
(30)
\[
\psi_{GCV}(\alpha, y^\delta_n) \leq \frac{1}{\rho(\alpha)} \left( C\alpha^{2\mu + 1} + \delta_n^2 \right), \quad \mu \leq \frac{1}{2}.
\]
(31)

**Proof.** It is a standard result [13] that \( ||T_n x^\delta - y^\delta_n - (T_n x_\alpha - y_n)|| \leq ||y^\delta_n - y_n|| \leq \delta_n \). Similarly, by the usual source condition, we obtain \( ||(T_n x_\alpha - y_n)|| \leq C\alpha^{2\mu + 1} \) for \( \mu \leq \frac{1}{2} \). The result follows from the triangle inequality. \( \square \)

The proceeding results generally follow from the infinite dimensional setting and we similarly obtain the following bounds from below:

**Lemma 4.** Suppose that \( \alpha \in I \) and also that (27) holds. Then
\[
\psi_{GCV}(\alpha, y^\delta_n - y_n) \geq \frac{1}{\rho(\alpha)} \left( C \frac{\eta^2}{\alpha^{2p-2}} \right).
\]
Moreover, if \( ||T_n x^\delta|| \geq C_0 \), with an \( n \)-independent constant, then there exists an \( n \)-independent constant \( C \) with
\[
\psi_{GCV}(\alpha, y_n) \geq C \frac{1}{\rho(\alpha)} \alpha^2.
\]

If (28) holds and \( \alpha \in I \), then
\[
\psi_{GCV}(\alpha, y_n) \geq C \frac{1}{\rho(\alpha)} \alpha^{2\mu + 1 + \varepsilon_2}, \quad \mu \leq \frac{1}{2}.
\]

**Theorem 4.** Let \( \mu \leq \frac{1}{2p} \) assume \( \alpha_* \) is the minimizer of \( \psi_{GCV}(\alpha, y_n^\delta) \) and suppose further that \( \alpha_* \in I \) such that (27) holds. Then
\[
\alpha_* \geq \left[ \inf_{\alpha \geq \alpha_*} \left( C\alpha^{2\mu + 1} + C\delta_n^2 \right) \right]^{-\frac{1}{2\mu + 1 - p - \varepsilon_2}} \geq C\delta_n^{\frac{\rho}{p - \varepsilon}} \eta_{\rho - \varepsilon}^{\frac{2}{p - \varepsilon}}.
\]
On the other hand
\[
\alpha_* \leq \left[ \inf_{\alpha \geq \alpha_*} \frac{1}{\rho(\alpha)} \left( C\alpha^{2\mu+1} + C\delta_n^2 \right) \right]^{1/2},
\]
with \( t = 2 \). If \( \alpha_* \in 1 \) and (28) hold, then the above upper bound on \( \alpha_* \) holds with \( t = 2\mu + 1 + \epsilon_2 \).

**Proof.** Take an arbitrary \( \bar{\alpha} \) and consider first the case \( \alpha_* \leq \bar{\alpha} \). Following on from the previous lemmas and using (30), we have
\[
\frac{1}{\rho(\alpha_*)} \left( C\frac{\eta^2}{\alpha_*^{2\mu+1 - \epsilon}} \right) \leq \psi_{\text{GCV}}^2(\alpha_*, y_n^\delta - y_n) \leq C\psi_{\text{GCV}}^2(\alpha_*, y_n^\delta) + C\psi_{\text{GCV}}^2(\alpha_*, y_n)
\]
\[
\leq \psi_{\text{GCV}}^2(\alpha^*, y_n^\delta) + C\frac{1}{\rho(\alpha_*)} \alpha_*^{2\mu+1} \leq \frac{1}{\rho(\bar{\alpha})} \left( C\alpha^{2\mu+1} + \delta_n^2 \right) + C\frac{1}{\rho(\alpha_*)} \alpha_*^{2\mu+1}.
\]

Hence, by the monotonicity of \( \alpha \mapsto \alpha^{2\mu+1} \) and since \( \rho \) is monotonically increasing, we obtain that
\[
\left( C\frac{\eta^2}{\alpha_*^{2\mu+1 - \epsilon}} \right) \leq \frac{\rho(\alpha_*)}{\rho(\bar{\alpha})} \left( C\alpha^{2\mu+1} + \delta_n^2 \right) + \alpha_*^{2\mu+1} \leq \left( C\alpha^{2\mu+1} + \delta_n^2 \right) + \alpha_*^{2\mu+1} \leq \left( C\alpha^{2\mu+1} + \delta_n^2 \right).
\]

Hence,
\[
\alpha_* \geq \left[ \inf_{\alpha \geq \alpha_*} \left( C\alpha^{2\mu+1} + C\delta_n^2 \right) \right]^{-\frac{1}{2\mu+1 - \epsilon}} \eta^{\frac{2\mu+1}{2\mu+1 - \epsilon}} \geq C\delta_n^{\frac{2\mu+1 - \epsilon}{2\mu+1}} \eta^{\frac{2\mu+1 - \epsilon}{2\mu+1}}.
\]

Now, suppose \( \alpha_* \geq \bar{\alpha} \). Then using that \( \alpha_* \) is a minimizer
\[
\frac{C}{\rho(\alpha_*)} \alpha_*^{2\mu+1} \leq \psi_{\text{GCV}}^2(\alpha_*, y_n) \leq C\psi_{\text{GCV}}^2(\alpha_*, y_n^\delta) + \psi_{\text{GCV}}^2(\alpha_*, y_n^\delta - y_n)
\]
\[
\leq \frac{1}{\rho(\bar{\alpha})} \left( C\alpha^{2\mu+1} + C\delta_n^2 \right) + C\frac{1}{\rho(\alpha_*)} \delta_n^2 \leq \frac{1}{\rho(\bar{\alpha})} \left( C\alpha^{2\mu+1} + C\delta_n^2 \right) + C\frac{1}{\rho(\alpha_*)} \delta_n^2.
\]

Hence, as \( \rho(\alpha_*) \) is bounded from above by 1, it follows that
\[
\alpha_* \leq \left[ \inf_{\alpha \leq \alpha_*} \frac{1}{\rho(\alpha)} \left( C\alpha^{2\mu+1} + C\delta_n^2 \right) \right]^{1/2}.
\]

**Theorem 5.** Suppose that \( \mu \leq \frac{1}{2}, \alpha_* \) is the minimizer of \( \psi_{\text{GCV}}(\alpha, y_n^\delta) \), where \( \alpha_* \in 1 \) such that (27) and (28) are satisfied. Suppose further that one has \( \rho(\delta_n^{2\mu+1}) \geq C \). Then
\[
\|x_{\alpha_*}^\delta - x^\dagger\| \leq \delta_n^{\frac{2\mu}{2\mu+1}} + \delta_n \left( \frac{\eta}{\delta_n} \right)^{\frac{1}{2\mu+1}},
\]
with \( t \) as in Theorem 4.
**Proof.** Since

\[ ||x_{z*}^\delta - x^\dagger|| \leq C x_{z*}^\mu + C \frac{\delta_n}{\sqrt{x_{z*}}}, \]

we may take the balancing parameter \( \bar{\alpha} = \frac{\delta_n}{\sqrt{x_{z*}}} \). From the previous theorem, it follows that if \( x_{z*} \leq \bar{\alpha} \), then

\[ x_{z*} \geq \frac{\eta^{2\mu+1}}{\inf_{x \geq x_\mu} (C x^{2\mu+1} + C \delta_n^2)} \geq \left( \frac{\eta}{\delta_n} \right)^{\frac{2}{\mu+1}}. \]

On the other hand, if \( x_{z*} \geq \bar{\alpha} \), and \( \rho(\bar{\alpha}) \geq C \), then

\[ x_{z*} \leq C \delta_n^2. \]

Thus, taking for \( x_{z*}^\mu \) and \( \frac{\delta_n}{\sqrt{x_{z*}}} \) the worst of these estimates, we obtain the desired result.

This result establishes convergence rates in the discrete case. However, the required conditions are somewhat restrictive as we need that the selected \( x_{z*} \) has to be in a certain interval (although this is to be expected in a finite-dimensional setting). Note that the term \( \delta_n^2 \) in Theorem 4 can be replaced by any reasonable monotonically decreasing upper bound for \( \psi_{GCV}(x, y_n^\delta - y_n) \). In particular, if we could conclude that \( x_{z*} \) is in a region where \( \psi_{GCV}(x, y_n^\delta - y_n) \leq C \frac{\eta}{\delta_n} \), then we would obtain similar convergence results as for the predictive mean square error.

Moreover, the condition that \( \rho(\delta_n^{2\mu+1}) > C \) restricts the analysis to the weakly bounded noise scenario, in which case \( \delta_n \to \infty \) as \( n \to \infty \). The standard “bounded noise” case is ruled out in Theorem 5, because if \( \delta_n \) would tend to zero, then this would lead to a contradiction with Proposition 2.

In general, however, the performance of the GCV-rule for the regularization of deterministic inverse problems is subpar compared to other heuristic rules, for example, those mentioned in the previous sections; cf., for example [23, 24]. This is also illustrated by the fact that we had to impose stronger conditions for the convergence results compared to the aforementioned rules.

### 4. Conclusion

We analyzed and provided conditions for the derivation of convergence rates for a number of well-known heuristic parameter choice rules in the weakly bounded noise setting and modified them when necessary. The theory was extended in a consistent and systematic way whereby one attains
the standard results whenever the situation is as in the classical setting. In particular, we provided noise conditions which are very often satisfied for when one can prove suboptimal convergence rates for the quasi-optimality, modified heuristic discrepancy and Hanke-Raus rules, as well as optimal rates whenever certain regularity conditions are satisfied.

A further novel aspect of this article was the examination of the generalized cross-validation rule and the predictive mean-square error in a deterministic framework. In the case of the former, it was in a finite-dimensional setting where we proved convergence rates under somewhat restrictive conditions. However, according to our analysis, the GCV rule does not present a preferable choice compared to the three aforementioned rules in the weakly bounded noise case.

In essence, it was demonstrated that heuristic rules remain viable methods for selecting the regularization parameter, even in the case where the noise is only weakly bounded.

Acknowledgment

The authors thank the reviewers for their valuable input which helped to improve the rigor as well as the presentation of the article.

Funding

This work was supported by the Austrian Science Fund (FWF) project P 30157-N31.

ORCID

Stefan Kindermann http://orcid.org/0000-0002-3688-5125
Kemal Raik http://orcid.org/0000-0002-3972-5565

References

[1] Jin, Q., Mathé, P. (2013). Oracle inequality for a statistical Raus–Gfrerer-type rule. SIAM/ASA J. Uncertainty Quantification 1(1):386–407. DOI: 10.1137/120889149.
[2] Lu, S., Mathé, P. (2014). Discrepancy based model selection in statistical inverse problems. J. Complexity 30(3):290–308. DOI: 10.1016/j.jco.2014.02.002.
[3] Burger, M., Helin, T., Kekkonen, H. (2016). Large noise in variational regularization. arXiv Prepr. 1602.00520. DOI: 10.1093/imatrm/tny002.
[4] Egger, H. (2008). Regularization of inverse problems with large noise. J. Phys: Conf. Ser. 124:012022. DOI: 10.1088/1742-6596/124/1/012022.
[5] Eggermont, P., LaRiccia, V., Nashed, M. (2009). On weakly bounded noise in ill-posed problems. Inverse Problems 25(11):115018. DOI: 10.1088/0266-5611/25/11/115018.
[6] Eggermont, P., LaRiccia, V., Nashed, M. (2012). On moment discretization for ill-posed problems with discrete weakly bounded noise. *Int. J. Geomath.* 3(2):155–178. DOI: 10.1007/s13137-012-0037-2.

[7] Eggermont, P., LaRiccia, V., Nashed, M. (2015). Noise models for ill-posed problems. *Handbk. Geomath.* 2nd ed. 1633–1658. DOI: 10.1007/978-3-642-54551-1.

[8] Kokurin, M. Y. (2006). Stable iteratively regularized gradient method for nonlinear irregular equations under large noise. *Inverse Problems* 22(1):197. DOI: 10.1088/0266-5611/22/1/011.

[9] Mathé, P., Tautenhahn, U. (2011). Regularization under general noise assumptions. *Inverse Problems* 27(3):035016. DOI: 10.1088/0266-5611/27/3/035016.

[10] Morozov, V. (1996). Regulyarizatsiya pri bol’shikh pomehah (Russian) (Regularization under large noise). *Zh. Vychisl. Mat. Mat. Fiz.* 36:13.

[11] Nashed, M. (1976). *Generalized Inverses and Applications: Proceedings of an Advanced Seminar*, Vol. 1973. Madison, WI: Publication of the Mathematics Research Center, The University of Wisconsin.

[12] Tikhonov, A., Glasko, V. (1964). The approximate solution of Fredholm integral equations of the first kind. *USSR Comput. Mathematics Math. Phys.* 4(3):236–247. DOI: 10.1016/0041-5553(64)90254-X.

[13] Engl, H., Hanke, M., Neubauer, A. (1996). *Regularization of Inverse Problems, Mathematics and its Applications*. Amsterdam, The Netherlands: Springer.

[14] Kindermann, S. (2011). Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems, electron. *Trans. Numer. Anal.* 38:233–257.

[15] Kindermann, S., Neubauer, A. (2008). On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization. *Inverse Problems Imaging* 2:291–299. DOI: 10.3934/ipi.2008.2.291.

[16] Hanke, M., Raus, T. (1996). A general heuristic for choosing the regularization parameter in ill-posed problems. *SIAM J. Sci. Comput.* 17(4):956–972. DOI: 10.1137/0917062.

[17] Raus, T., Hämärik, U. (2018). Heuristic parameter choice in Tikhonov method from minimizers of the quasi-optimality function. In: *New Trends in Parameter Identification for Mathematical Models, Trends Mathematics*, Cham: Birkhäuser/ Springer, pp. 227–244.

[18] Bakushinskii, A. (1985). Remarks on choosing a regularization parameter using quasi-optimality and ratio criterion. *USSR Comput. Math. Math. Phys.* 24:181–182.

[19] Kindermann, S., Pereverzyev, S., Jr., Pilipenko, A. (2018). The quasi-optimality criterion in the linear functional strategy. *Inverse Problems* 34(7):075001–075024. DOI: 10.1088/1361-6420/aabef4.

[20] Lukas, M. A. (1993). Asymptotic optimality of generalized cross-validation for choosing the regularization parameter. *Numer. Math.* 66(1):41–66. DOI: 10.1007/BF01385687.

[21] Wahba, G. (1990). *Spline Models for Observational Data*, Vol. 59 of CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

[22] Fenu, C., Reichel, L., Rodriguez, G., Sadok, H. (2017). GCV for tikhonov regularization by partial SVD. *Bit Numer. Math.* 57(4):1019–1039. DOI: 10.1007/s10543-017-0662-0.
[23] Bauer, F., Lukas, M. A. (2011). Comparing parameter choice methods for regularization of ill-posed problems. *Math. Comput. Simul.* 81(9):1795–1841. DOI: 10.1016/j.matcom.2011.01.016.

[24] Hämarik, U., Palm, R., Raus, T. (2011). Comparison of parameter choices in regularization algorithms in case of different information about noise level. *Calcolo.* 48:47–59. DOI: 10.1007/s10092-010-0027-4.