Electronic Transactions on Numerical Analysis.
Volume 40, pp. 58-81, 2013.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

DISCRETIZATION INDEPENDENT CONVERGENCE RATES FOR NOISE LEVEL-FREE PARAMETER CHOICE RULES FOR THE REGULARIZATION OF ILL-CONDITIONED PROBLEMS*

STEFAN KINDERMANN†

Abstract. We develop a convergence theory for noise level-free parameter choice rules for Tikhonov regularization of finite-dimensional, linear, ill-conditioned problems. In particular, we derive convergence rates with bounds that do not depend on troublesome parameters such as the small singular values of the system matrix. The convergence analysis is based on specific qualitative assumptions on the noise, the noise conditions, and on certain regularity conditions. Furthermore, we derive several sufficient noise conditions both in the discrete and infinite-dimensional cases. This leads to important conclusions for the actual implementation of such rules in practice. For instance, we show that for the case of random noise, the regularization parameter can be found by minimizing a parameter choice functional over a subinterval of the spectrum (whose size depends on the smoothing properties of the forward operator), yielding discretization independent convergence rate estimates, which are of optimal order under regularity assumptions for the exact solution.

Key words. regularization, parameter choice rule, Hanke-Raus rule, quasioptimality rule, generalized cross validation

AMS subject classifications. 65J20, 47A52, 65J22

1. Introduction.

The selection of the regularization parameter for regularization methods is of high importance for ill-posed problems. Several methods for this task are known; see, e.g., [5]. It is standard to use parameter choice rules which depend on the knowledge of the noise level. Here instead, we discuss so-called noise level-free (or heuristic, or data-driven) parameter choice rules, which do not use the noise level at all, for instance, the quasioptimality principle, the Hanke-Raus rules, and generalized cross validation. For a long time, such methods have been considered of minor importance since it is well-known that they cannot yield convergence in the worst case for ill-posed problems. However, recently [10, 11, 15] a successful, quite general theory has been developed for such rules within the framework of a restricted noise analysis (using so-called noise conditions). It is the aim of this paper to extend this theory to the case of discrete ill-conditioned problems. Such problems typically (but not exclusively) arise by discretizing ill-posed problems and are very important. The transfer of the convergence theory from the infinite-dimensional case to the finite-dimensional one is not as straightforward as might be expected at first look. The reason is that the standard noise conditions as formulated in [10, 11, 15] are never satisfied in the discrete case. A central topic in this paper is to replace these conditions by ones that are useful in a finite-dimensional setting. The aim is to develop a convergence theory proving convergence rates that are robust with respect to the discretization, i.e., estimates with constants that do not depend on "bad" parameters such as the condition number, which usually blows up as the discretization becomes finer.

We establish such a convergence theory by imitating the infinite-dimensional theory and using appropriate noise conditions and regularity conditions. Our analysis is based on Tikhonov regularization but can be extended to other methods as well.

The paper is organized as follows: In Section 2 we set the stage, define the parameter choice rules that we consider, and state the main abstract convergence theorem. The next section, Section 3, is an extension of the study of the noise and regularity conditions in [10].

*Received March 28, 2012. Accepted August 5, 2012. Published online on April 12, 2013. Recommended by L. Reichel.

†Industrial Mathematics Institute, Johannes Kepler University of Linz, Altenbergerstr. 69, 4040 Linz, Austria (kindermann@indmath.uni-linz.ac.at).

ETNA
Kent State University
http://etna.math.kent.edu

We develop alternative formulations of these in the infinite-dimensional case. This section is rather independent of the other ones and puts previous results (such as the noise conditions in [11] and scaling conditions of [2]) onto a common ground. In Section 4 we state analogous noise and regularity conditions which are meaningful in the discrete case. The main difference to the infinite-dimensional case is that the relevant conditions only have to hold in a subinterval of the spectrum away from 0. Using this, we find convergence rates for the quasioptimality principle and the Hanke-Raus rules, similar to the infinite-dimensional case (Theorem 4.8). Furthermore, we also show some rate estimates for generalized cross validation. In Section 5, we study the noise conditions under the assumption that the noise is white (independent and identically normally distributed) and estimate the probability that they are satisfied, which immediately leads in combination with Theorem 4.8 to convergence rates which hold with a certain probability. By studying some typical examples of ill-conditioned problems in Section 5.1, we can consider the question if and in what sense a noise condition holds, and how the parameter choice rules are used in practice. This is elaborated on in the final section.

2. Tikhonov regularization and parameter choice functionals. In the following, we focus on ill-conditioned linear equations in Hilbert spaces

$$A_n x_n = y_{\delta,n}, \tag{2.1}$$

where A_n is an operator $X_n \to Y_n$ with finite-dimensional range, x_n is the unknown solution, and $y_{\delta,n}$ are given (possibly noisy) data. Note that we always assume that (2.1) is ill-conditioned (i.e., A_n has a large condition number), hence in this paper we use the terms “discrete” and “ill-conditioned” as synonyms. Below we will prove results which also hold in the infinite-dimensional case, e.g., in Section 3. To indicate this situation, we will drop the subscript n and write A, x, y, X, Y, etc., when all involved Hilbert spaces are allowed to be infinite-dimensional as well.

Formally, the problem (2.1) is never ill-posed. Nevertheless ill-conditioned problems show in many cases a similar behavior as ill-posed ones and usually have to be approached by regularization. Indeed, suppose that the data $y_{\delta,n}$ in (2.1) are a contamination of some exact data y_n by noise with noise level δ,

$$y_{\delta,n} = y_n + e_n, \qquad \|e_n\| = \delta.$$

Then solving (2.1) by means of the Moore-Penrose pseudo-inverse $x_{n,\delta} = A_n^\dagger y_{\delta,n}$ leads to an error bound

$$\|x_n^\dagger - x_{n,\delta}\| \leq \frac{1}{\sigma_{min}(A_n)}\delta, \tag{2.2}$$

where $\sigma_{min}(A_n)$ is the smallest singular value of A_n and $x_n^\dagger$ denotes the (unknown) solution for exact data $x_n^\dagger = A_n^\dagger y_n$. For discretized versions of ill-posed problems, this bound is sharp and becomes very large as the discretization becomes finer, such that quite often $x_{n,\delta}$ is of no use at all. The remedy in this situation is to use regularization, for instance, Tikhonov regularization, and to compute

$$x_{\alpha,\delta,n} = \left(A_n^T A_n + \alpha I\right)^{-1} A_n^T y_{\delta,n}, \tag{2.3}$$

where $\alpha > 0$ is the regularization parameter.

Usually, the regularized solution $x_{\alpha,\delta,n}$ is a better estimate of $x_n^\dagger$ than $x_{n,\delta}$ if the regularization parameter is chosen appropriately. Several methods for this choice are well-known

ETNA
Kent State University
http://etna.math.kent.edu

such as a priori or a posteriori rules, which require the knowledge of the noise level; see, e.g., [5]. Recently, a convergence theory was developed for noise level-free parameter choice rules, which have the advantage that the noise level is not needed. However, the price to pay is that they only work if the data error e_n satisfies some additional *noise conditions* [10, 11, 15]. It is the aim of this article to extend the analysis of [10] to the case of discrete, ill-conditioned problems (2.1).

The class of noise level-free rules we are considering here has the following form. The regularization parameter $\alpha = \alpha^*$ is selected by minimizing a certain functional ψ

$$\alpha^* = \text{argmin}_{\alpha \in I_n} \psi(\alpha, y_{\delta,n}) \tag{2.4}$$

over a compact interval $I_n \subset [0, \infty]$. Note that in the discrete case, the interval I_n of possible regularization parameters has to exclude 0, i.e., $I_n \subset [\eta, \infty[$ with $\eta > 0$, contrary to the infinite-dimensional case, where an interval $[0, \alpha_0]$ can be chosen. This is an subtle but important issue and will be addressed in more detail in Sections 4 and 5, where examples of possible intervals I_n are given.

The main goal of our analysis is to prove, if possible, *discretization independent* error estimates, i.e., estimates of the form

$$\|x_{\alpha,\delta,n} - x_n^\dagger\| \leq f(\delta),$$

where f is robust in the sense that it stays bounded even if A_n approaches an operator A of an infinite-dimensional ill-posed problem. In particular, f should not depend on constants such as the smaller singular values of A_n. Note that (2.2) is not discretization independent, since the smallest singular value usually tends to 0 as n increases if A_n represents a discretization of a forward operator A of an ill-posed problem.

In this paper we do *not* consider convergence as $n \to \infty$, in particular, we do not assume that A_n converges to some operator A with infinite-dimensional range. Nevertheless, the results of this paper are still relevant in several situations: it can be the case that the problem (2.1) is given without reference to any discretization of a infinite-dimensional problem, such that it is only of interest to compute the solution $x_n^\dagger$ in a stable way. Discretization independent error estimates are of great use because they can be applied no matter how high the condition number of A_n is. The second case is that A_n is a suitable discretization of an ill-posed problem with an operator A, which has the property that $x_n^\dagger$ converges to the solution $x^\dagger$ of the infinite-dimensional problem. This happens, for instance, when using the dual projection method [5]. In this case, it is not difficult to use our results to prove convergence of $x_{\alpha,\delta,n}$ as the discretization becomes finer, $n \to \infty$. However, it is well-known that not every discretization A_n of such A leads to convergence of $x_n^\dagger$ to $x^\dagger$; see the Seidman counterexample [17]. In this case, our estimates are useless, but this does not necessarily mean that heuristic parameter choice rules cannot be applied successfully.

We will focus on Tikhonov regularization, but the analysis can be extended to other methods as well [10]. Furthermore, we will use the following notations: by an index function we mean a continuous, strictly increasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\phi(0) = 0$. We will denote by y_n the exact data and by $x_n^\dagger$ the associated minimum norm solution,

$$y_n = A_n x_n^\dagger, \qquad x_n^\dagger = A_n^\dagger y_n,$$

$y_{\delta,n}$ will denote the noisy data

$$y_{\delta,n} = y_n + e_n,$$

ETNA
Kent State University
http://etna.math.kent.edu

with δ the noise level

$$\delta = \|y_{\delta,n} - y_n\| = \|e_n\|.$$

Moreover, $x_{\alpha,\delta,n}$ will denote the regularized solution (2.3), while $x_{\alpha,n}$ is the regularized solution with exact data

$$x_{\alpha,n} = \left(A_n^T A_n + \alpha I\right)^{-1} A_n^T y_n.$$

As it is standard, the total error $\|x_{\alpha,\delta,n} - x_n^\dagger\|$ is split into two parts, the propagated data error

$$e_d(\alpha) = \|x_{\alpha,\delta,n} - x_{\alpha,n}\|$$

and the approximation error

$$e_a(\alpha) = \|x_{\alpha,n} - x_n^\dagger\|,$$

which yields the well-known error bound

$$\|x_{\alpha,\delta,n} - x_n^\dagger\| \leq e_d(\alpha) + e_a(\alpha). \tag{2.5}$$

For a fixed parameter choice functional, the regularization parameter α^* is selected via (2.4). The main estimate of the total error is then given in the following theorem; see also [10, 11, 15].

THEOREM 2.1. *Let $\psi : I_n \times Y \to \mathbb{R}$ be subadditive, i.e.,*

$$\psi(\alpha, z + w) \leq \psi(\alpha, z) + \psi(\alpha, w),$$

nonnegative, $\psi(\alpha) \geq 0$, and symmetric, $\psi(\alpha, -y) = \psi(\alpha, y)$, and let a minimum α^* in (2.4) exist.

For any fixed $e_n \in Y_n, z_n \in R(A_n)$, let there exist a monotonically decreasing function $\rho_{\downarrow,e_n} : \mathbb{R}^+ \to \mathbb{R}^+$ and a monotonically increasing function $\rho_{\uparrow,z_n} : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\psi(\alpha, e_n) \leq \rho_{\downarrow,e_n}(\alpha) \quad \forall \alpha \in I_n, \tag{2.6}$$

$$\psi(\alpha, z_n) \leq \rho_{\uparrow,z_n}(\alpha) \quad \forall \alpha \in I_n. \tag{2.7}$$

Moreover, for $y_n = A_n x_n^\dagger$ fixed, let there exists a set $\mathcal{N} \subset Y_n$ (the set of "admissible noise"), a constant C_d, and an index functions Φ such that

$$e_d(\alpha^*) \leq C_d \psi(\alpha^*, y_{\delta,n} - y_n) \quad \forall y_{\delta,n} - y_n \in \mathcal{N}, \tag{2.8}$$

$$e_a(\alpha^*) \leq \Phi(\psi(\alpha^*, y_n)). \tag{2.9}$$

Then, if $y_{\delta,n} - y_n \in \mathcal{N}$ we have the error estimate

$$\|x_{\alpha^*,\delta,n} - x_n^\dagger\| \leq \inf_{\alpha \in I_n} \begin{cases} C_d \rho_{\downarrow,y_{\delta,n}-y_n}(\alpha) + 2C_d \rho_{\uparrow,y_n}(\alpha) + e_a(\alpha) & \text{if } \alpha^* \leq \alpha, \\ \Phi\left(\rho_{\downarrow,y_{\delta,n}-y_n}(\alpha) + 2\rho_{\uparrow,y_n}(\alpha)\right) + e_d(\alpha) & \text{if } \alpha^* > \alpha. \end{cases} \tag{2.10}$$

Proof. The proof is an adaption of those in [10, 15]. Note that $e_d(\alpha)$ is monotonically decreasing and $e_a(\alpha)$ is monotonically increasing. Let $\alpha \in I_n$ be arbitrary but fixed. If $\alpha^* \leq \alpha$, then by monotonicity it holds that $e_a(\alpha^*) \leq e_a(\alpha)$. By the properties and estimates of ψ and the monotonicity of $\rho_{\uparrow,y}$, we find

$$\begin{aligned} e_d(\alpha^*) &\leq C_d \psi(\alpha^*, y_{\delta,n} - y_n) \leq C_d \left(\psi(\alpha^*, y_{\delta,n}) + \psi(\alpha^*, y_n)\right) \\ &\leq C_d \psi(\alpha, y_{\delta,n}) + C_d \rho_{\uparrow,y_n}(\alpha^*) \leq C_d \left(\rho_{\downarrow,y_{\delta,n}-y_n}(\alpha) + \rho_{\uparrow,y_n}(\alpha)\right) + C_d \rho_{\uparrow,y_n}(\alpha) \\ &\leq C_d \rho_{\downarrow,y_{\delta,n}-y_n}(\alpha) + 2C_d \rho_{\uparrow,y_n}(\alpha). \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

With (2.5) the result follows for the case $\alpha^* \leq \alpha$. Now assume that $\alpha^* > \alpha$. Then the role of e_d and e_a are reversed:

$$\begin{aligned} &e_d(\alpha^*) \leq e_d(\alpha), \\ &e_a(\alpha^*) \leq \Phi\left(\psi(\alpha^*, y_{\delta,n}) + \psi(\alpha^*, y_{\delta,n} - y_n)\right) \leq \Phi\left(\psi(\alpha, y_{\delta,n}) + \rho_{\downarrow, y_{\delta,n}-y_n}(\alpha^*)\right) \\ &\qquad \leq \Phi\left(\rho_{\uparrow, y_n}(\alpha) + \rho_{\downarrow, y_{\delta,n}-y_n}(\alpha) + \rho_{\downarrow, y_{\delta,n}-y_n}(\alpha)\right). \qquad \square \end{aligned}$$

REMARK 2.2. The functions $\rho_{\uparrow,y_n}$ and $\rho_{\downarrow,y_{\delta,n}-y_n}$ are usually known and independent of n. Calculating the infimum in (2.10) then leads to discretization independent convergence (rates) if all the prerequisites of this theorem hold with n-independent constants. Here, the most difficult part is to show (2.8) and (2.9). The estimate on the noise term, (2.8), is in general impossible without restricting the noise to a set $\mathcal{N}$. The condition that $y_{\delta,n} - y_n \in \mathcal{N}$ such that (2.8) holds, is termed a *noise condition* and the bound (2.10) is a bound in the *restricted noise case*. The corresponding inequality (2.9) is a condition on the exact solution. Note that in general the desirable choice $\Phi(x) = C_a x$ does not hold with a discretization independent constant, but only $\Phi(x) = C_a x^\nu$ with $\nu < 1$. Because of this, the right-hand side of (2.10) only yields suboptimal rates. However, if so-called *regularity conditions* (also called decay conditions in [10, 11]) on the exact solution hold, then an estimate (2.9) with $\Phi(x) = C_a x$ and a discretization independent constant C_a is possible, leading to optimal order estimates.

REMARK 2.3. This theorem is neither specific to the discrete case nor to Tikhonov regularization. It remains valid in the infinite-dimensional case and also for other *monotone* regularization methods; see [10] for details.

2.1. Parameter choice functionals and their bounds. We are now in the position to analyze specific parameter choice functionals and the corresponding estimates in Theorem 2.1 in more detail. For the analysis, the parameter choice functionals are conveniently expressed in terms of a spectral family of the operator A. Since we are focusing on the discrete case, we only need to consider the singular value decomposition of A_n.

Let us denote by $(\sigma_i, u_i, v_i)_{i=1}^N$ the singular system (with σ_i the singular values, u_i the left and v_i the right singular vectors) of the operator A_n with finite-dimensional range. We consider the following rules obtained via (2.4) and the corresponding parameter functionals: the *quasioptimality* rule [19, 20], where $\psi = \psi_{QO}$, the *Hanke-Raus rule HR*$_1$ [9], where $\psi = \psi_{HR,1}$, the *Hanke-Raus rule HR*$_\infty$ [9], where $\psi = \psi_{HR,\infty}$, and *generalized cross validation* [21], where $\psi = \psi_{GCV}$. These functionals are defined as follows, using $\lambda_i := \sigma_i^2$.

$$\begin{aligned} \psi_{QO}(\alpha, y_{\delta,n})^2 &= \sum_{i=1}^N \frac{\alpha^2 \lambda_i}{(\lambda_i + \alpha)^4} |(y_{\delta,n}, v_i)|^2, \\ \psi_{HR,1}(\alpha, y_{\delta,n})^2 &= \sum_{i=1}^N \frac{\alpha^2}{(\lambda_i + \alpha)^3} |(y_{\delta,n}, v_i)|^2, \\ \psi_{HR,\infty}(\alpha, y_{\delta,n})^2 &= \sum_{i=1}^N \frac{\alpha}{(\lambda_i + \alpha)^2} |(y_{\delta,n}, v_i)|^2, \\ \psi_{GCV}(\alpha, y_{\delta,n})^2 &= \frac{1}{\left(\frac{1}{N}\sum_{i=1}^N \frac{\alpha}{\alpha+\lambda_i}\right)^2} \sum_{i=1}^N \left(\frac{\alpha}{(\lambda_i + \alpha)}\right)^2 |(y_{\delta,n}, v_i)|^2. \end{aligned}$$

For each of these rules, the regularization parameter is chosen by minimizing the functional with respect to α as in (2.4) using only the actual given data $y_{\delta,n}$ as information.

ETNA
Kent State University
http://etna.math.kent.edu

Note that for the computation of these functionals and α^*, the singular system is not needed. For instance, the quasioptimality rule is in practice computed by selecting a sequence of geometrically decreasing regularization parameters, $\alpha_i = \alpha_0 q^i \subset I_n$, with $q < 1$, and choosing $\alpha^* = \alpha_i$, where i is the integer where the minimum of

$$\|x_{\alpha_{i+1},\delta,n} - x_{\alpha_i,\delta,n}\| \qquad i = 1, 2, \ldots$$

is attained. More precisely, this is the discrete quasioptimality rule [19], but the corresponding functional can be treated quite similar to the original one. The rule HR_1 can be computed by

$$\psi_{HR,1}(\alpha, y_{\delta,n})^2 = \alpha^{-1}(y_{\delta,n} - A_n x^{II}_{\alpha,\delta,n}, y_{\delta,n} - A_n x_{\alpha,\delta,n}),$$

employing one step of the iterated Tikhonov regularization [9]

$$x^{II}_{\alpha,\delta,n} := x_{\alpha,\delta,n} + (A_n^T A_n + \alpha I)^{-1}\left(A_n^T(y_{\delta,n} - A_n x_{\alpha,\delta,n})\right).$$

The rule HR_∞ is particularly simple, since it is just an appropriately α-scaled residual

$$\psi_{HR,\infty}(\alpha, y_{\delta,n})^2 = \alpha^{-1}\|A_n x_{\alpha,\delta,n} - y_{\delta,n}\|^2,$$

and in a similar way, the GCV-functional can be computed by

$$\psi_{GCV}(\alpha, y_{\delta,n})^2 = \eta(\alpha)^2\|A_n x_{\alpha,\delta} - y_{\delta,n}\|^2 \tag{2.11}$$

with

$$\eta(\alpha) = \left(\frac{\alpha}{N}\text{trace}\left((A_n^* A_n + \alpha I)^{-1}\right)\right)^{-1}.$$

For more information, further functionals, and possible fine-tuning, we refer to [3, 6, 7, 10, 16]. All of the above defined parameter choice functionals are obviously positive, symmetric, and subadditive, and by continuity, a minimum α^* in (2.4) always exists. In view of Theorem 2.1 we are now interested in the corresponding estimates.

The propagated data error and the approximation error can be expressed in terms of the spectral system as

$$e_d(\alpha)^2 = \sum_{i=1}^N \frac{\lambda_i}{(\alpha+\lambda_i)^2}|(y_{\delta,n} - y_n, v_i)|^2, \qquad e_a(\alpha)^2 = \sum_{i=1}^N \frac{\alpha^2}{(\alpha+\lambda_i)^2}|(x_n^\dagger, u_i)|^2,$$

which immediately yields the following estimates of type (2.6), (2.7):

$$\begin{aligned}
\psi_{QO}(\alpha, y_{\delta,n} - y_n) &\le e_d(\alpha), \\
\psi_{QO}(\alpha, y_n) &\le e_a(\alpha), \\
\psi_{HR,1}(\alpha, y_{\delta,n} - y_n) &\le \frac{\delta}{\sqrt{\alpha}}, \\
\psi_{HR,1}(\alpha, y_n) &\le e_a(\alpha), \\
\psi_{HR,\infty}(\alpha, y_{\delta,n} - y_n) &\le \frac{\delta}{\sqrt{\alpha}}, \qquad (2.12)\\
\psi_{HR,\infty}(\alpha, y_n) &\le c_\nu \alpha^\nu \|x_n^\dagger\|_{-\nu} \quad \forall 0 < \nu \le \frac{1}{2}. \qquad (2.13)
\end{aligned}$$

Here $\|x_n^\dagger\|_{-\nu}$ is the norm of a "source element" in a source condition

$$\|x_n^\dagger\|_{-\nu}^2 = \|(A_n^T A_n)^{-\nu} x_n^\dagger\|^2 = \sum_{i=1}^N \frac{1}{\lambda_i^{2\nu}}|(x^\dagger, u_i)|^2, \tag{2.14}$$

and $c_\nu^2 = \frac{(1+2\nu)^{1+2\nu}(1-2\nu)^{1-2\nu}}{4}$. Concerning ψ_{GCV}, we notice that it differs from $\psi_{HR,\infty}$ only by a function depending on α. Thus, we can use the bounds for $\psi_{HR,\infty}$ to get

$$\psi_{GCV}(\alpha, y_{\delta,n} - y_n) \leq \eta(\alpha)\delta$$
$$\psi_{GCV}(\alpha, y_n) \leq c_\nu \|x_n^\dagger\|_{-\nu}\eta(\alpha)\alpha^{\nu+\frac{1}{2}}, \quad \forall 0 < \nu \leq \frac{1}{2}.$$

Furthermore, we notice that estimates for $\psi_{HR,1}(\alpha, y_\delta - y)$ and $\psi_{HR,\infty}(\alpha, y_{\delta,n} - y_n)$ in terms of $e_d(\alpha)$ are also possible if additional (rather restrictive) conditions on the noise hold; see, e.g., [10, Lemma 4.9].

3. Noise conditions and regularity conditions. We now study the so-called noise conditions, i.e., conditions on $y_{\delta,n} - y_n$ such that inequalities of the form (2.8) hold. It was shown in [11] that such estimates are possible with bounded constants for the quasi-optimality principle, and later for many other combinations of regularization methods and parameter choice functionals [10].

At first we extend some known results on the noise condition using inequalities equivalent to or sufficient for (2.8). We will derive these in the general infinite-dimensional case, extending the analysis of [10, 11]; of course, the discrete setting is a special case of this. The infinite-dimensional versions of the parameter choice functionals $\psi_{QO}, \psi_{HR,1}, \psi_{HR,\infty}$ are obvious and can be found for instance in [10].

LEMMA 3.1. *Let $A : X \to Y$ be a bounded operator between Hilbert spaces, and let the regularization be Tikhonov regularization. The inequality (2.8) for $\psi = \psi_{QO}$ is equivalent to*

$$\int_1^\infty \psi_{QO}(\alpha^*\eta, y_\delta - y)^2 \frac{\eta - 1}{\eta^2}\, d\eta \leq \frac{C_d^2}{6}\psi_{QO}(\alpha^*, y_\delta - y)^2, \tag{3.1}$$

and for the case $\psi = \psi_{HR,1}$, the inequality (2.8) is equivalent to

$$\int_1^\infty \psi_{HR,1}(\alpha^*\eta, y_\delta - y)^2 \frac{\eta - 2}{\eta^2}\, d\eta \leq \frac{C_d^2}{2}\psi_{HR,1}(\alpha^*, y_\delta - y)^2. \tag{3.2}$$

Proof. Let F_λ be a spectral family of AA^*. For the quasioptimality functional ψ_{QO} we use Fubini's theorem to get

$$\int_1^\infty \psi_{QO}(\alpha^*\eta, y_\delta - y)^2 \frac{\eta - 1}{\eta^2}\, d\eta = \int_1^\infty \int_\sigma \frac{{\alpha^*}^2\lambda\eta^2}{(\lambda + \alpha^*\eta)^4}\, dF_\lambda \|y_\delta - y\|^2 \frac{\eta - 1}{\eta^2}\, d\eta$$
$$= \int_\sigma \int_1^\infty \frac{{\alpha^*}^2\lambda\eta^2}{(\lambda + \alpha^*\eta)^4} \frac{\eta - 1}{\eta^2}\, d\eta\, dF_\lambda \|y_\delta - y\|^2 = \frac{1}{6}\int_\sigma \frac{\lambda}{(\alpha^* + \lambda)^2}\, dF_\lambda \|y_\delta - y\|^2.$$

For $\psi = \psi_{HR,1}$ we find similarly using Q, the orthogonal projector onto $\overline{R(A)}$,

$$\int_1^\infty \psi_{HR,1}(\alpha^*\eta, y_\delta - y)^2 \frac{\eta - 2}{\eta^2}\, d\eta = \int_1^\infty \int_\sigma \frac{{\alpha^*}^2\eta^2}{(\lambda + \alpha^*\eta)^3}\, dF_\lambda \|Q(y_\delta - y)\|^2 \frac{\eta - 2}{\eta^2}\, d\eta$$
$$= \int_\sigma \int_1^\infty \frac{{\alpha^*}^2\eta^2}{(\lambda + \alpha^*\eta)^3} \frac{\eta - 2}{\eta^2}\, d\eta\, dF_\lambda \|Q(y_\delta - y)\|^2 = \frac{1}{2}\int_\sigma \frac{\lambda}{(\alpha^* + \lambda)^2}\, dF_\lambda \|y_\delta - y\|^2.$$

Since $e_d(\alpha^*)^2 = \int_\sigma \frac{\lambda}{(\alpha^*+\lambda)^2}\, dF_\lambda \|y_\delta - y\|^2$, the assertion follows. ∎

From this lemma we can find several sufficient conditions such that (2.8) holds. We define

$$V(t) := \int_0^t \lambda\, dF_\lambda \|y_\delta - y\|^2, \qquad W(t) := \int_0^t dF_\lambda \|Q(y_\delta - y)\|^2.$$

ETNA
Kent State University
http://etna.math.kent.edu

PROPOSITION 3.2. *Let the same assumptions as in Lemma 3.1 hold. For $\psi = \psi_{QO}$, each of the following conditions imply (3.1) and hence (2.8).*

- *There exists an $\epsilon > 0$ and $C_d < \infty$ such that*

$$\psi_{QO}(\alpha^*\eta, y_\delta - y)^2 \leq \epsilon(1+\epsilon)\frac{C_d^2}{6}\eta^{-\epsilon}\psi_{QO}(\alpha^*, y_\delta - y)^2, \quad \forall \eta \geq 1. \tag{3.3}$$

- *There exists an $\epsilon > 0$ and $C_d < \infty$ such that*

$$\int_1^\infty V(\eta t)\frac{\eta - 1}{\eta^4}\, d\eta \leq \frac{C_d^2}{6} V(t) \quad \forall t > 0. \tag{3.4}$$

- *There exists an $\epsilon > 0$ and $C_d < \infty$ such that*

$$V(\eta t) \leq \epsilon(1+\epsilon)\frac{C_d^2}{6}\eta^{2-\epsilon}V(t) \quad \forall \eta \geq 1, t > 0. \tag{3.5}$$

- *There exists a constant $C_{nc} < \infty$ such that*

$$\int_t^\infty \frac{1}{\lambda}\, dF_\lambda \|y_\delta - y\|^2 \leq \frac{C_{nc}}{t^2}\int_0^t \lambda\, dF_\lambda\|y_\delta - y\|^2 \quad \forall t > 0. \tag{3.6}$$

Here, (3.6) holds if and only if (3.5) holds.

Proof. The first condition (3.3) implies (3.1) simply by integration. For (3.4) we use integration by parts and a change of variables

$$\begin{aligned}
\psi_{QO}(\alpha^*, y_\delta - y)^2 &= \int_0^\infty \frac{{\alpha^*}^2\lambda}{(\alpha^* + \lambda)^4}\, dF_\lambda\|y_\delta - y\|^2 \\
&= \int_0^\infty \frac{{\alpha^*}^2}{(\alpha^*+\lambda)^4}\, dV(\lambda) = 4\int_0^\infty \frac{{\alpha^*}^2}{(\alpha^*+\lambda)^5} V(\lambda)\, d\lambda, \\
\psi_{QO}(\alpha^*\eta, y_\delta - y)^2 &= 4\frac{1}{\eta^2}\int_0^\infty \frac{{\alpha^*}^2}{(\alpha^*+\xi)^5}V(\xi\eta)\, d\xi,
\end{aligned}$$

from which the sufficiency of (3.4) for (3.1) follows. Again by integration, (3.5) implies (3.4). The equivalence of (3.6) and (3.5) is a consequence of a celebrated theorem of Arinjo and Muckenhoupt [1]; see [18]. The constant C_{nc} in (3.6) can be related to C_d by inspection of the proofs in [11]. □

In particular, the noise condition of Theorem 2.1 can be formulated by defining $\mathcal{N}$ as the set of all $y_\delta - y$ such that one of the conditions in this proposition holds (with discretization independent constants). For the Hanke-Raus rule HR_1 we have similar characterizations.

PROPOSITION 3.3. *Let the same assumptions as in Lemma 3.1 hold. For $\psi = \psi_{HR,1}$, each of the following conditions imply (3.2) and hence (2.8).*

- *There exist an $\epsilon > 0$ and $C_d < \infty$ such that*

$$\psi_{HR,1}(\alpha^*\eta, y_\delta - y)^2 \leq 2^\epsilon\epsilon(1+\epsilon)\frac{C_d^2}{2}\eta^{-\epsilon}\psi_{HR,1}(\alpha^*, y_\delta - y)^2, \quad \forall \eta \geq 2. \tag{3.7}$$

- *There exist an $\epsilon > 0$ and $C_d < \infty$ such that*

$$\int_2^\infty W(\eta t)\frac{\eta - 2}{\eta^3}\, d\eta \leq \frac{C_d^2}{2}W(t) \quad \forall t > 0. \tag{3.8}$$

ETNA
Kent State University
http://etna.math.kent.edu

- *There exist an* $\epsilon > 0$ *and* $C_d < \infty$ *such that*

$$W(\eta t) \leq 2^{\epsilon}\epsilon(1+\epsilon)\frac{C_d^2}{2}\eta^{1-\epsilon}W(t) \quad \forall \eta \geq 2, \forall t > 0. \tag{3.9}$$

- *There exists a constant* $C_{nc} > 0$ *such that*

$$\int_t^{\infty} \frac{1}{\lambda}\, dF_\lambda \|Q(y_\delta - y)\|^2 \leq \frac{C_{nc}}{t} \int_0^t dF_\lambda \|Q(y_\delta - y)\|^2 \quad \forall t > 0. \tag{3.10}$$

Here, (3.10) *holds if and only if* (3.9) *holds.*

Proof. The implication (3.7) $\Rightarrow$ (3.2) follows by multiplication of (3.2) by $\frac{\eta-2}{\eta^2}$ and integration over $\eta > 2$, noting that the part of the integral in (3.2) over $\eta \in [1, 2]$ is negative. In view of (3.8), we can proceed as for the quasioptimality case,

$$\psi_{HR,1}(\alpha^*, y_\delta - y)^2 = \int_0^{\infty} \frac{{\alpha^*}^2}{(\alpha^* + \lambda)^3}\, dW(\lambda) = 3\int_0^{\infty} \frac{{\alpha^*}^2}{(\alpha^* + \lambda)^4} W(\lambda)\, d\lambda,$$

$$\psi_{HR,1}(\alpha^*\eta, y_\delta - y)^2 = 3\frac{1}{\eta}\int_0^{\infty} \frac{{\alpha^*}^2}{(\alpha^* + \xi)^4} W(\eta\xi)\, d\xi,$$

which implies (3.2) after integration. By integration over η, (3.9) implies (3.8). Finally, again by the results of [1], (3.10) is equivalent to (3.9), when the condition holds for all $\eta > 1$. But it is straightforward to see that this is also equivalent to the condition holding for all $\eta > 2$, possibly with different constants. □

To complete the picture we recall results of [10] for the Hanke-Raus rule HR_∞:

PROPOSITION 3.4. *If* (3.10) *or* (3.9) *holds, then* (2.8) *holds (possibly with a different constant* C_d *than in* (3.9)*) for* $\psi_{HR,\infty}$.

The conditions (3.10) and (3.6) were used in [10, 11] to establish (2.8) for several regularization methods. Our analysis shows the sufficiency of the scaling-type conditions. This type of conditions (but stronger ones) were employed in [2] to prove convergence rates for the quasioptimality principle.

The noise conditions are usually interpreted as restrictions that rule out "smooth noise", i.e., noise that is in the range of A. This can be seen in the following proposition. Here we denote again by Q the orthogonal projector onto $\overline{R(A)}$.

PROPOSITION 3.5. *If* $Q(y_\delta - y) \neq 0$ *and if one of the conditions* (3.4), (3.5), (3.6), (3.8), (3.9), *or* (3.10) *holds, then* $Q(y_\delta - y) \notin R(A)$. *In particular, if* A *has finite-dimensional range, then none of these conditions can hold.*

Proof. Since (3.5) or (3.6) imply (3.4), and (3.9) or (3.10) imply (3.8), it is enough to prove this proposition if either (3.4) or (3.8) holds. Suppose that $Q(y_\delta - y) \in R(A)$. Then

$$W(t) \leq o(t) \text{ and } V(t) \leq o(t^2).$$

In this case, assuming (3.4), we find by the change of variables $z = \eta t$ that

$$t^2 \int_t^{\infty} \frac{V(z)}{z^4}(z - t)\, dz = t^2 \int_0^{\infty} \frac{V(z)}{z^4} \max\{z - t, 0\}\, dz \leq o(t^2).$$

The function $(z, t) \mapsto \max\{z - t, 0\}$ is nonnegative and monotonically increasing as $t \to 0$. By the monotone convergence theorem, we obtain that

$$\int_0^{\infty} \frac{V(z)}{z^3}\, dz = \lim_{t\to 0} \int_0^{\infty} \frac{V(z)}{z^4} \max\{z - t, 0\}\, dz = 0,$$

ETNA
Kent State University
http://etna.math.kent.edu

which is impossible unless $V(z) = 0$ almost everywhere. Using (3.8) we can conclude analogously the absurd consequence that

$$\int_0^\infty \frac{W(z)}{z^3}\,dz = \lim_{t\to 0}\int_0^\infty \frac{W(z)}{z^3}\max\{z-2t,0\}\,dz = 0.$$

Hence, $Q(y_\delta - y) \notin R(A)$. Clearly, this can only hold for nonzero $Q(y_\delta - y)$ if $\overline{R(A)} \neq R(A)$, i.e., only when A has non-closed range and hence never in the discrete or well-posed case. □

In the following sections we will consider noise conditions that make sense in the discrete case.

3.1. Regularity conditions. Besides the noise condition, the estimate (2.9) is the second main ingredient in Theorem 2.1. The situation here is different to that in (2.8) because (2.9) is already satisfied for some index function if a source condition holds [10]. Unfortunately, this only yields suboptimal rates. Of particular interest is the case when (2.9) holds with $\Phi(x) \sim x$, as this implies optimal order rates. Sufficient conditions for this situation were stated in [11] and were called decay conditions. Here, we will use the term regularity condition instead. Thus, we are now interested in finding properties of $x^\dagger$ that allows us to conclude that

$$e_a(\alpha^*) \leq C_a \psi(\alpha^*, Ax^\dagger) \tag{3.11}$$

holds for some of the parameter choice functionals. To begin with, we again study the infinite-dimensional case extending previous results of [10, 11]. The following is an analogue of Lemma 3.1.

LEMMA 3.6. *Let $A : X \to Y$ be a bounded operator between Hilbert spaces, and let the regularization be Tikhonov regularization. The inequality* (3.11) *for $\psi = \psi_{QO}$ is equivalent to*

$$\int_1^\infty \psi_{QO}(\frac{\alpha^*}{\eta}, Ax^\dagger)^2 \frac{\eta - 1}{\eta^2}\,d\eta \leq \frac{C_a^2}{6}\psi_{QO}(\alpha^*, Ax^\dagger)^2, \tag{3.12}$$

and for the case $\psi = \psi_{HR,1}$, the inequality (3.11) *is equivalent to*

$$\int_1^\infty \psi_{HR,1}(\frac{\alpha^*}{\eta}, y_\delta - y)^2 \frac{1}{\eta}\,d\eta \leq \frac{C_a^2}{2}\psi_{HR,1}(\alpha^*, y_\delta - y)^2. \tag{3.13}$$

Proof. Denote by E_λ a spectral family of A^*A. The approximation error can be expressed as $e_a(\alpha) = \int \frac{\alpha^2}{(\alpha+\lambda)^2}\,dE_\lambda \|x^\dagger\|^2$. Hence, the lemma follows from

$$\int_1^\infty \frac{(\frac{\alpha^*}{\eta})^2\lambda^2}{(\frac{\alpha^*}{\eta}+\lambda)^4}\frac{\eta-1}{\eta^2}\,d\eta = \frac{1}{6}\frac{\alpha^{*2}}{(\alpha^*+\lambda)^2}, \qquad \int_1^\infty \frac{(\frac{\alpha^*}{\eta})^2\lambda}{(\frac{\alpha^*}{\eta}+\lambda)^3}\frac{1}{\eta}\,d\eta = \frac{1}{2}\frac{\alpha^{*2}}{(\alpha^*+\lambda)^2}.$$ □

From this we may derive sufficient conditions for (3.11) in form of scaling conditions. Let us define

$$\tilde{V}(t) = \int_t^\infty \frac{1}{\lambda^2}\,dE_\lambda \|x^\dagger\|^2.$$

PROPOSITION 3.7. *Let the same assumptions as in Lemma 3.6 hold. Each of the following conditions imply* (3.12) *(and hence* (3.11) *for the quasioptimality rule).*

ETNA
Kent State University
http://etna.math.kent.edu

- *There exist an* $\epsilon > 0$ *and* $C_a < \infty$ *such that*

$$\psi_{QO}(\frac{\alpha^*}{\eta}, Ax^\dagger)^2 \leq \epsilon(1+\epsilon)\frac{C_a^2}{6}\eta^{-\epsilon}\psi_{QO}(\alpha^*, Ax^\dagger)^2 \quad \forall \eta \geq 1. \tag{3.14}$$

- *There exist an* $\epsilon > 0$ *and* $C_a < \infty$ *such that*

$$\int_1^\infty \tilde{V}(\frac{t}{\eta})\frac{\eta - 1}{\eta^4}\, d\eta \leq \frac{C_a^2}{6}\tilde{V}(t) \quad \forall t \geq 0. \tag{3.15}$$

- *There exist an* $\epsilon > 0$ *and* $C_a < \infty$ *such that*

$$\tilde{V}(\frac{t}{\eta}) \leq \epsilon(1+\epsilon)\frac{C_a^2}{6}\tilde{V}(t)\eta^{2-\epsilon} \quad \forall \eta \geq 1, t > 0. \tag{3.16}$$

- *There exist constants* C_{rc}, t_1 *such that*

$$\int_0^t dE_\lambda \|x^\dagger\|^2 \leq C_{rc}t^2 \int_t^\infty \frac{1}{\lambda^2}\, dE_\lambda \|x^\dagger\|^2 \quad \forall 0 \leq t \leq t_1. \tag{3.17}$$

Moreover, each of the following conditions imply (3.13) *(and hence* (3.11)*) for the Hanke-Raus rule* HR_1.

- *There exist an* $\epsilon > 0$ *and* $C_a < \infty$ *such that*

$$\psi_{HR,1}(\frac{\alpha^*}{\eta}, Ax^\dagger)^2 \leq \epsilon\frac{C_a^2}{2}\eta^{-\epsilon}\psi_{HR,1}(\alpha^*, Ax^\dagger)^2 \quad \forall \eta \geq 1. \tag{3.18}$$

- *There exist an* $\epsilon > 0$ *and* $C_a < \infty$ *such that*

$$\int_1^\infty \tilde{V}(\frac{t}{\eta})\frac{1}{\eta^3}\, d\eta \leq \frac{C_a^2}{2}\tilde{V}(t) \quad \forall t \geq 0. \tag{3.19}$$

- *There exist an* $\epsilon > 0$ *and* $C_a < \infty$ *such that*

$$\tilde{V}(\frac{t}{\eta}) \leq \epsilon\frac{C_a^2}{2}\tilde{V}(t)\eta^{2-\epsilon} \qquad \forall \eta \geq 1, t > 0. \tag{3.20}$$

- *Condition* (3.17).

Proof. The conditions (3.14) and (3.18) imply (3.12) and (3.13), respectively, by integration. By an integration by parts we find that

$$\psi_{QO}(\alpha^*, Ax^\dagger)^2 = 4\int_0^\infty \frac{{\alpha^*}^3\lambda^3}{(\alpha^* + \lambda)^5}\tilde{V}(\lambda)\, d\lambda,$$

$$\psi_{HR,1}(\alpha^*, Ax^\dagger)^2 = 3\int_0^\infty \frac{{\alpha^*}^3\lambda^2}{(\alpha^* + \lambda)^4}\tilde{V}(\lambda)\, d\lambda,$$

which shows that (3.15) and (3.19) imply (3.12) and (3.13), respectively. By integration, (3.16) and (3.20) imply the corresponding inequalities (3.15) and (3.19). The sufficiency of (3.17) was already shown in [11, 15]. ∎

Concerning the HR_∞ rule, the estimate (3.11) was established under (3.17) in [10]. We remark that regularity conditions of the form (3.17) have already been used in [10, 11]. It should be noticed that if a source condition with saturation index holds, i.e.,

$$\int_0^\infty \frac{1}{\lambda^2}\, dE_\lambda \|x^\dagger\|^2 < \infty,$$

then (3.17) is automatically satisfied; see [10, 11].

ETNA
Kent State University
http://etna.math.kent.edu

4. Discrete case. Proposition 3.5 indicates a difficulty that occurs in the discrete case. Since then $Q(y_{\delta,n} - y_n)$ is always in the range of A_n ($A_n^\dagger$ is defined on the whole space), the noise conditions as mentioned in Proposition 3.5 cannot be satisfied. This also can be observed by a limit argument: in all cases $\psi(\alpha, y_{\delta,n} - y_n)$ tends to 0 as $\alpha \to 0$ while $\lim_{\alpha\to 0} e_d(\alpha) = A_n^\dagger(y_{\delta,n} - y_n)$. More precisely, we have

LEMMA 4.1. *If A_n has finite-dimensional range, then for all $z \in Y_n$ the functionals $\psi_{QO}(\alpha, z)$ and $\psi_{HR,\infty}(\alpha, z)$ are monotonically increasing in α for $\alpha \in [0, \sigma_{min}^2]$ with $\psi_{QO}(0, z) = 0$, $\psi_{HR,\infty}(0, z) = 0$, and $\psi_{HR,1}(\alpha, z)$ is monotonically increasing in α for $\alpha \in [0, 2\sigma_{min}^2]$ with $\psi_{HR,1}(0, z) = 0$.*

Thus, the estimate (2.8) cannot be satisfied uniformly for all α^* sufficiently small. This is the reason why in the discrete case one has to restrict the search for a minimum of ψ to an interval which does not contain 0. The following propositions are appropriate formulations of noise conditions in the discrete case. They are analogous to (3.6) and (3.10).

PROPOSITION 4.2. *Let us define*

$$e_i = (y_{\delta,n} - y_n, v_i).$$

If

$$\inf_{\tau\geq 0}\left\{(1+\tau)^2 + \frac{\sum_{\lambda_i>\tau\alpha^*}\frac{e_i^2}{\lambda_i}}{\sum_{\lambda_i\leq\tau\alpha^*}\lambda_i e_i^2}{\alpha^*}^2(1+\tau)^4\right\} \leq C_d^2,$$

then (2.8) holds for ψ_{QO}. If

$$\inf_{\tau>0}\left\{\tau(1+\tau) + (1+\tau)^3\alpha^*\frac{\sum_{\lambda_i>\tau\alpha^*}\frac{e_i^2}{\lambda_i}}{\sum_{\lambda_i\leq\tau\alpha}e_i^2}\right\} \leq C_d^2,$$

then (2.8) holds for $\psi_{HR,1}$. If

$$\inf_{\tau>0}\left\{\tau + (1+\tau)^2\alpha^*\frac{\sum_{\lambda_i>\tau\alpha^*}\frac{e_i^2}{\lambda_i}}{\sum_{\lambda_i\leq\tau\alpha^*}e_i^2}\right\} \leq C_d^2,$$

then (2.8) holds for $\psi_{HR,\infty}$.

Proof. Let $\tau > 0$ be arbitrary. Then

$$\begin{aligned}
e_d(\alpha)^2 &= \sum_{\lambda_i\leq\tau\alpha^*}\frac{\lambda_i}{(\alpha^*+\lambda_i)^2}e_i^2 + \sum_{\lambda_i>\tau\alpha^*}\frac{\lambda_i}{(\alpha^*+\lambda_i)^2}e_i^2\\
&\leq (1+\tau)^2\sum_{\lambda_i\leq\tau\alpha^*}\frac{{\alpha^*}^2\lambda_i}{(\alpha^*+\lambda_i)^4}e_i^2 + \sum_{\lambda_i>\tau\alpha^*}\frac{1}{\lambda_i}e_i^2\\
&\leq (1+\tau)^2\psi_{QO}(\alpha^*, y_\delta - y)^2\\
&\quad + \frac{\sum_{\lambda_i>\tau\alpha^*}\frac{1}{\lambda_i}|(y_\delta - y, v_i)|^2}{\sum_{\lambda_i\leq\tau\alpha^*}\lambda_i|(y_\delta - y, v_i)|^2}{\alpha^*}^2(1+\tau)^4\psi_{QO}(\alpha^*, y_\delta - y)^2,
\end{aligned}$$

where the last inequality follows from

$$\sum_{\lambda_i\leq\tau\alpha^*}\lambda_i e_i^2 \leq {\alpha^*}^2(1+\tau)^4\sum_{\lambda_i\leq\tau\alpha^*}\frac{{\alpha^*}^2\lambda_i}{(\alpha^*+\lambda_i)^4}e_i^2.$$

ETNA
Kent State University
http://etna.math.kent.edu

In a similar fashion we obtain

$$\sum_{\lambda_i \leq \tau\alpha^*} \frac{\lambda_i}{(\alpha^*+\lambda_i)^2} e_i^2$$
$$\leq \begin{cases} \tau \sum_{\lambda_i \leq \tau\alpha^*} \frac{\alpha^*}{(\alpha^*+\lambda_i)^2} e_i^2 \leq \tau \psi_{HR,\infty}(\alpha^*, y_\delta - y)^2, \\ \tau(1+\tau) \sum_{\lambda_i \leq \tau\alpha^*} \frac{\alpha^{*2}}{(\alpha^*+\lambda_i)^3} e_i^2 \leq \tau(1+\tau)\psi_{HR,1}(\alpha^*, y_\delta - y)^2, \end{cases}$$
$$\sum_{\lambda_i \leq \tau\alpha^*} e_i^2$$
$$\leq \begin{cases} \alpha^*(1+\tau)^2 \sum_{\lambda_i \leq \tau\alpha^*} \frac{\alpha^*}{(\alpha^*+\lambda_i)^2} e_i^2 \leq \alpha^*(1+\tau)^2 \psi_{HR,\infty}(\alpha^*, y_\delta - y)^2, \\ \alpha^*(1+\tau)^3 \sum_{\lambda_i \leq \tau\alpha^*} \frac{\alpha^{*2}}{(\alpha^*+\lambda_i)^3} e_i^2 \leq \alpha^*(1+\tau)^3\psi_{HR,1}(\alpha^*, y_\delta - y)^2. \end{cases} \qquad \square$$

As a simple consequence we obtain a discrete version of the noise condition allowing α^* to be in an interval:

PROPOSITION 4.3. *In the case of the quasioptimality rule, $\psi = \psi_{QO}$, let the following condition hold: there exists a constant C_{ncd} and an interval $I_n \subset [0,\infty)$ such that*

$$\xi^2 \sum_{\lambda_i > \xi} \frac{e_i^2}{\lambda_i} \leq C_{ncd} \sum_{\lambda_i \leq \xi} \lambda_i e_i^2 \quad \forall \xi \in I_n. \tag{4.1}$$

Then, for any $\tau > 0$, the noise condition (2.8) holds for all $\alpha^* \in \frac{1}{\tau} I_n$ with a constant

$$C_d^2 = (1+\tau)^2 + C_{ncd}\frac{(1+\tau)^4}{\tau^2}\,.$$

In the case of the Hanke-Raus rules, $\psi = \psi_{HR,1}$ or $\psi = \psi_{HR,\infty}$, let the following condition hold: there exists a constant C_{ncd} and an interval $I_n \subset [0,\infty)$ such that

$$\xi \sum_{\lambda_i > \xi} \frac{e_i^2}{\lambda_i} \leq C_{ncd} \sum_{\lambda_i \leq \xi} e_i^2 \quad \forall \xi \in I_n. \tag{4.2}$$

Then for any $\tau > 0$, the noise condition (2.8) holds for all $\alpha^* \in \frac{1}{\tau} I_n$ with a constant

$$C_d^2 = \tau(1+\tau) + C_{ncd}\frac{(1+\tau)^3}{\tau},$$

in the case of $\psi = \psi_{HR,1}$ and with a constant

$$C_d^2 = \tau + C_{ncd}\frac{(1+\tau)^2}{\tau},$$

in the case of $\psi = \psi_{HR,\infty}$.

Proof. A proof follows by setting $\xi = \tau\alpha^*$. $\square$

In Section 5 we will look closer at the conditions (4.1), (4.2) for the case of random noise. Let us now consider the regularity (decay) condition in the discrete case. First, we are interested in estimates of the form

$$e_a(\alpha^*) \leq C_a \psi(\alpha^*, Ax_n^\dagger)^\nu, \quad 0 < \nu \leq 1, \tag{4.3}$$

for some ν. The most important case, $\nu = 1$, which yields optimal order rates, will be treated in Proposition 4.6. We recall the definition of $\|.\|_{-\nu}$ in (2.14).

ETNA
Kent State University
http://etna.math.kent.edu

PROPOSITION 4.4. *Let $0 < \nu \leq 1$ be fixed. If*

$$(\nu^\nu(1-\nu)^{(1-\nu)})^2 \|x_n^\dagger\|_{-\nu}^2 \inf_{\eta>0} \frac{(\alpha^*+\eta)^{4\nu}}{\eta^{4\nu}\left(\sum_{\lambda_i\geq\eta}\lambda^{-2}|(x_n^\dagger,u_i)|^2\right)^\nu} \leq C_a^2, \tag{4.4}$$

then (4.3) *is satisfied for ψ_{QO}. If*

$$(\nu^\nu(1-\nu)^{(1-\nu)})^2 \|x_n^\dagger\|_{-\nu}^2 \inf_{\eta>0} \frac{(\alpha^*+\eta)^{3\nu}}{\eta^{3\nu}\left(\sum_{\lambda_i\geq\eta}\lambda^{-2}|(x_n^\dagger,u_i)|^2\right)^\nu} \leq C_a^2, \tag{4.5}$$

then (4.3) *is satisfied for $\psi_{HR,1}$ and $\psi_{HR,\infty}$.*

Proof. A standard convergence rate estimate yields

$$e_a(\alpha^*)^2 \leq \sup_{x\in\mathbb{R}} \frac{x^{2\nu}}{(1+x)^2}\alpha^{*2\nu}\|x_n^\dagger\|_{-\nu}^2 \leq (\nu^\nu(1-\nu)^{(1-\nu)})^2\|x_n^\dagger\|_{-\nu}^2\alpha^{*2\nu}.$$

For arbitrary $\eta > 0$ we have

$$\begin{aligned}\psi_{QO}(\alpha^*,Ax_n^\dagger)^2 &= \sum_i \frac{\alpha^{*2}\lambda_i^2}{(\alpha^*+\lambda_i)^4}|(x_n^\dagger,u_i)|^2 \geq \alpha^{*2}\sum_i \frac{\lambda_i^4}{(\alpha^*+\lambda_i)^4}\frac{(x_n^\dagger,u_i)|^2}{\lambda_i^2} \\ &\geq \alpha^{*2}\frac{\eta^4}{(\alpha^*+\eta)^4}\sum_{\lambda_i\geq\eta}\frac{(x_n^\dagger,u_i)|^2}{\lambda_i^2},\end{aligned}$$

which proves (4.4). For $\psi_{HR,1}$ and for $\psi_{HR,\infty}$ we have

$$\begin{aligned}\psi_{HR,1}(\alpha^*,Ax_n^\dagger)^2 &= \sum_i \frac{\alpha^{*2}\lambda_i}{(\alpha^*+\lambda_i)^3}|(x_n^\dagger,u_i)|^2 \geq \alpha^{*2}\sum_i \frac{\lambda_i^3}{(\alpha^*+\lambda_i)^3}\frac{(x_n^\dagger,u_i)|^2}{\lambda_i^2} \\ &\geq \alpha^{*2}\frac{\eta^3}{(\alpha^*+\eta)^3}\sum_{\lambda_i\geq\eta}\frac{(x_n^\dagger,u_i)|^2}{\lambda_i^2}, \\ \psi_{HR,\infty}(\alpha^*,Ax_n^\dagger)^2 &= \sum_i \frac{\alpha^{*}\lambda_i}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2 \geq \alpha^{*2}\sum_i \frac{\lambda_i^3}{\alpha^*(\alpha^*+\lambda_i)^2}\frac{(x_n^\dagger,u_i)|^2}{\lambda_i^2} \\ &\geq \alpha^{*2}\frac{\eta^3}{(\alpha^*+\eta)^3}\sum_{\lambda_i\geq\eta}\frac{(x_n^\dagger,u_i)|^2}{\lambda_i^2},\end{aligned}$$

which yields (4.5). □

The relevance of this proposition is that basically only a discretization independent source condition is enough to obtain (4.3) with uniform constants. More precisely,

COROLLARY 4.5. *If $\|x_n^\dagger\|_{-\nu} \leq c_1$ and for some η*

$$\sum_{\lambda_i\geq\eta}\lambda^{-2}|(x_n^\dagger,u_i)|^2 \geq c_2, \tag{4.6}$$

then for all $\alpha^* \leq [0,\alpha_{max}]$, (4.3) *is satisfied for $\psi_{QO},\psi_{HR,1},\psi_{HR,\infty}$ with such a ν and a constant $C_a = \nu^\nu(1-\nu)^{1-\nu}c_1\left(\frac{(\alpha_{max}+\eta)^\omega}{\eta^\omega c_2}\right)^{\frac{\nu}{2}}$, where $\omega = 4$ for ψ_{QO} and $\omega = 3$ for $\psi_{HR,1},\psi_{HR,\infty}$.*

The condition (4.6) is not difficult to satisfy. It only means that the low frequency part of $x_n^\dagger$ does not become too small as the discretization becomes finer. When we want to

ETNA
Kent State University
http://etna.math.kent.edu

have (4.3) with $\nu = 1$, we can apply the previous proposition only when a source condition at the saturation holds, i.e., when $\|x_n^\dagger\|_{-1}$ is bounded by a discretization independent constant. However, even if this is not the case, we can use the following regularity conditions and are still able to satisfy (4.3) with $\nu = 1$:

PROPOSITION 4.6. *If*

$$\inf_\tau \frac{(\tau+1)^2}{\tau^2}\left(\frac{(\tau+1)^2}{\tau^2{\alpha^*}^2}\frac{\sum_{\lambda_i<\alpha^*}|(x_n^\dagger,u_i)|^2}{\sum_{\lambda_i\geq\tau\alpha^*}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2}}+1\right)\leq C_a^2, \tag{4.7}$$

then (4.3) *is satisfied for* ψ_{QO} *with* $\nu = 1$. *If*

$$\inf_\tau \frac{(\tau+1)}{\tau}\left(\frac{(\tau+1)^2}{\tau^2{\alpha^*}^2}\frac{\sum_{\lambda_i<\tau\alpha^*}|(x_n^\dagger,u_i)|^2}{\sum_{\lambda_i\geq\tau\alpha^*}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2}}+1\right)\leq C_a^2, \tag{4.8}$$

then (4.3) *is satisfied for* $\psi_{HR,1}$ *with* $\nu = 1$. *If*

$$\inf_\tau \frac{1}{\tau}\left(\frac{(\tau+1)^2}{\tau^2{\alpha^*}^2}\frac{\sum_{\lambda_i<\tau\alpha^*}|(x_n^\dagger,u_i)|^2}{\sum_{\lambda_i\geq\tau\alpha^*}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2}}+1\right)\leq C_a^2, \tag{4.9}$$

then (4.3) *is satisfied for* $\psi_{HR,\infty}$ *with* $\nu = 1$

Proof. We estimate the functionals ψ from below as above

$$\sum_i \frac{{\alpha^*}^2\lambda_i^2}{(\alpha^*+\lambda_i)^4}|(x_n^\dagger,u_i)|^2 \geq \frac{\tau^2}{(1+\tau)^2}\sum_{\lambda_i\geq\tau\alpha^*}\frac{{\alpha^*}^2}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2,$$

$$\sum_i \frac{{\alpha^*}^2\lambda_i}{(\alpha^*+\lambda_i)^3}|(x_n^\dagger,u_i)|^2 \geq \frac{\tau}{(1+\tau)}\sum_{\lambda_i\geq\tau\alpha^*}\frac{{\alpha^*}^2}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2,$$

$$\sum_i \frac{\alpha^*\lambda_i}{(\alpha^*+\lambda_i)2}|(x_n^\dagger,u_i)|^2 \geq \tau\sum_{\lambda_i\geq\tau\alpha^*}\frac{{\alpha^*}^2}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2,$$

and

$$\sum_{\lambda_i\geq\tau\alpha^*}\frac{{\alpha^*}^2}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2 \geq \frac{\tau^2}{(1+\tau)^2}{\alpha^*}^2\sum_{\lambda_i\geq\tau\alpha^*}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2}.$$

Moreover,

$$\begin{aligned}
e_a(\alpha^*)^2 &\leq \sum_{\lambda_i<\tau\alpha^*}|(x_n^\dagger,u_i)|^2 + \sum_{\lambda_i\geq\tau\alpha^*}\frac{{\alpha^*}^2}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2 \\
&\leq \frac{1}{{\alpha^*}^2}\frac{\sum_{\lambda_i<\tau\alpha^*}|(x_n^\dagger,u_i)|^2}{\sum_{\lambda_i\geq\tau\alpha^*}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2}}{\alpha^*}^2\sum_{\lambda_i\geq\tau\alpha^*}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2} + \sum_{\lambda_i\geq\tau\alpha^*}\frac{{\alpha^*}^2}{(\alpha^*+\lambda_i)^2}|(x_n^\dagger,u_i)|^2.
\end{aligned}$$

Combining the inequalities yields the proof. □

As a corollary we have the following result:

COROLLARY 4.7. *If there exists a constant* C_{rcd} *such that*

$$\sum_{\lambda_i\leq\xi}|(x_n^\dagger,u_i)|^2 \leq C_{rcd}\,\xi^2\sum_{\lambda_i\geq\xi}\frac{|(x_n^\dagger,u_i)|^2}{\lambda_i^2} \qquad \forall\xi\in I_n, \tag{4.10}$$

ETNA
Kent State University
http://etna.math.kent.edu

then (4.7), (4.8), (4.9) *are satisfied (and thus* (3.11) *with* $\nu = 1$*) for all* $\alpha^* \in \frac{1}{\tau} I_n$ *with constants* $C_a^2 = \{\frac{(\tau+1)^2}{\tau^2}, \frac{(\tau+1)}{\tau}, \frac{1}{\tau}\}((\tau+1)^2 C_{rcd}+1)$ *for* ψ_{QO}, $\psi_{HR,1}$, *and* $\psi_{HR,\infty}$, *respectively.*

We can collect the results into the main theorem which states sufficient conditions to obtain discretization independent estimates.

THEOREM 4.8. *Let* $\nu > 0$, *define*

$$\tilde{\Psi}_\nu(x) := \max\{x^\nu, x\},$$

and let $x_n^\dagger$ *satisfy either*

1. *a source condition* $\|x_n^\dagger\|_{-\nu} \leq c_1$ *and* (4.6), *or*
2. *a regularity condition* (4.10) *on an interval* I_n.

Then, if the parameter choice is the quasioptimality principle and (4.1) *is satisfied for the noise* $y_{\delta,n} - y_n$ *on the interval* I_n, *we obtain the convergence rate estimates*

$$\|x_{\alpha,\delta,n} - x_n^\dagger\| \leq C \inf_{\alpha \in I_n} \begin{cases} \tilde{\Psi}_\nu\left(\|x_{\alpha,n} - x_{\alpha,\delta,n}\| + \|x_{\alpha,n} - x_n^\dagger\|\right) & \textit{Case 1,} \\ \|x_{\alpha,n} - x_{\alpha,\delta,n}\| + \|x_{\alpha,n} - x_n^\dagger\| & \textit{Case 2.} \end{cases} \tag{4.11}$$

If the parameter choice is the rule HR$_1$, $\psi = \psi_{HR,1}$, *and* (4.2) *is satisfied on the interval* I_n, *then we obtain the convergence rate estimates*

$$\|x_{\alpha,\delta,n} - x_n^\dagger\| \leq C \inf_{\alpha \in I_n} \begin{cases} \tilde{\Psi}_\nu\left(\frac{\delta}{\sqrt{\alpha}} + \|x_{\alpha,n} - x_n^\dagger\|\right) & \textit{Case 1,} \\ \frac{\delta}{\sqrt{\alpha}} + \|x_{\alpha,n} - x_n^\dagger\| & \textit{Case 2.} \end{cases} \tag{4.12}$$

For the HR$_\infty$ *rule,* $\psi = \psi_{HR,\infty}$, *if* (4.2) *is satisfied and a source condition holds in either case, with* $\nu \leq \frac{1}{2}$, *we obtain the convergence rate estimate with* c_ν *as in* (2.13)

$$\|x_{\alpha,\delta,n} - x_n^\dagger\| \leq C \inf_{\alpha \in I_n} \begin{cases} \tilde{\Psi}_\nu\left(\frac{\delta}{\sqrt{\alpha}} + c_\nu \|x_n^\dagger\|_{-\nu} \alpha^\nu\right) & \textit{Case 1,} \\ \frac{\delta}{\sqrt{\alpha}} + c_\nu \|x_n^\dagger\|_{-\nu} \alpha^\nu & \textit{Case 2.} \end{cases} \tag{4.13}$$

The constants C *can be chosen as* $C = \max\{2^\nu C_a + 1, 2C_d + 1\}$, *with* C_d *as in Proposition 4.3 and* C_a *as in Corollary 4.5 or Corollary 4.7 with* $\tau = 1$.

Slight improvements of these results are possible using the proofs of [10]. For instance, with an additional condition on the noise, $\psi_{HR,1}$ satisfies (4.11); see [10, (4.19)].

In this theorem, we may allow (4.10) and (4.1) or (4.2) to hold only on a scaled interval $I_n' = \tau I_n$ with $\tau > 0$. In this case the constants C depend on τ by the expressions stated in Proposition 4.3, Corollary 4.5, and Corollary 4.7.

Using standard bounds, it is easy to see that (4.11) is a better estimate than (4.12), which is better than (4.13). Moreover, it should be noted that we only find optimal rates if the noise level is such that the optimal choice of the parameter (in general this is close to that α which balances the two terms in (4.11)–(4.13)) is in the interval I_n. This requires the noise conditions (4.1), (4.2) to hold on a sufficiently large interval, which is, unfortunately, not always the case; see below.

4.1. Generalized cross validation. We have established convergence rates for the quasioptimality and Hanke-Raus rules. Let us now discuss briefly the generalized cross validation. We remind of the similarity (2.11) of the GCV-functional with the HR$_\infty$ functional up to a factor depending on α. We have the following result.

ETNA
Kent State University
http://etna.math.kent.edu

PROPOSITION 4.9. *Let the assumptions of Theorem 4.8 for the Hanke-Raus functional $\psi_{HR,\infty}$ hold. Let α^* be chosen by the GCV-functional according to (2.4). Then we have the estimate*

$$\|x_{\alpha,\delta} - x_n^\dagger\| \leq C \inf_{\alpha \in I_n} \tilde{\Psi}_\nu(1 + \sqrt{\max\{\frac{\alpha^*}{\alpha}, \frac{\alpha}{\alpha^*}\}}) \begin{cases} \tilde{\Psi}_\nu\left(\frac{\delta}{\sqrt{\alpha}} + c_\nu \|x_n^\dagger\|_{-\nu}\alpha^\nu\right) & \textit{Case 1,} \\ \frac{\delta}{\sqrt{\alpha}} + c_\nu \|x_n^\dagger\|_{-\nu}\alpha^\nu & \textit{Case 2.} \end{cases}$$

Proof. Consider the factor

$$\frac{\psi_{GCV}(\alpha, y_\delta)}{\psi_{HR,\infty}(\alpha, y_\delta)} =: \rho(\alpha) = \sqrt{\alpha}\eta(\alpha) = \frac{1}{\alpha^{\frac{1}{2}}\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i + \alpha}\right)}.$$

Under the assumptions of Theorem 4.8, (3.11) and (2.8) are satisfied for $\psi_{HR,\infty}$. We proceed as in Theorem 2.1. Let $\alpha \in I_n$ be arbitrary and consider first the case $\alpha^* \leq \alpha$. Using (2.4), subadditivity, (2.13), and (2.12), we obtain with a generic constant C,

$$\begin{aligned} \|x_{\alpha^*,\delta,n} - x_n^\dagger\| &\leq e_a(\alpha^*) + e_d(\alpha^*) \leq e_a(\alpha) + \frac{C_d}{\rho(\alpha^*)}\psi_{GCV}(\alpha^*, y_\delta - y) \\ &\leq e_a(\alpha) + \frac{C_d}{\rho(\alpha^*)}\left(\psi_{GCV}(\alpha^*, y_\delta) + \psi_{GCV}(\alpha^*, y)\right) \\ &\leq e_a(\alpha) + \frac{C_d}{\rho(\alpha^*)}\psi_{GCV}(\alpha, y_\delta) + C_d C(\alpha^*)^\nu \\ &\leq (C_d + 1)C(\alpha)^\nu + \frac{C_d\rho(\alpha)}{\rho(\alpha^*)}\left(\frac{\delta}{\sqrt{\alpha}} + C(\alpha)^\nu\right) \\ &\leq C\left(1 + \frac{\rho(\alpha)}{\rho(\alpha^*)}\right)\left(\frac{\delta}{\sqrt{\alpha}} + C(\alpha)^\nu\right). \end{aligned}$$

In the case $\alpha^* > \alpha$, we obtain similarly,

$$\begin{aligned} \|x_{\alpha^*,\delta,n} - x_n^\dagger\| &\leq e_d(\alpha) + \frac{C_a}{\rho(\alpha^*)}\psi_{GCV}(\alpha^*, y) \\ &\leq \frac{\delta}{\sqrt{\alpha}} + C_a\left(\frac{\delta}{\sqrt{\alpha^*}} + \frac{\psi_{GCV}(\alpha, y_\delta)}{\rho(\alpha^*)}\right)^\nu \\ &\leq C_a\frac{\delta}{\sqrt{\alpha}} + C_a\left(\frac{\delta}{\sqrt{\alpha}} + \frac{\rho(\alpha)}{\rho(\alpha^*)}\left(\frac{\delta}{\sqrt{\alpha}} + C(\alpha)^\nu\right)\right)^\nu \\ &\leq C\left(1 + \frac{\rho(\alpha)}{\rho(\alpha^*)}\right)^\nu \tilde{\Psi}_\nu\left(\frac{\delta}{\sqrt{\alpha}} + C(\alpha)^\nu\right). \end{aligned}$$

Now we bound the factor $\frac{\rho(\alpha)}{\rho(\alpha^*)}$. It can be verified that

$$\frac{\alpha}{\alpha + \lambda_i} \geq \begin{cases} \frac{\alpha}{\alpha^* + \lambda_i} & \text{if } \alpha \leq \alpha^*, \\ \frac{\alpha^*}{\alpha^* + \lambda_i} & \text{if } \alpha > \alpha^*. \end{cases}$$

Thus,

$$\begin{aligned} \rho(\alpha)^2 &= \frac{\alpha}{\left(\frac{1}{n}\sum_{i=1}^n \frac{\alpha}{\alpha + \lambda_i}\right)^2} \leq \frac{\alpha}{\min\{\alpha, \alpha^*\}^2}\frac{1}{\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\alpha^* + \lambda_i}\right)^2} \\ &= \frac{\alpha\alpha^*}{\min\{\alpha, \alpha^*\}^2}\rho(\alpha^*)^2 = \max\{\frac{\alpha^*}{\alpha}, \frac{\alpha}{\alpha^*}\}\rho(\alpha^*)^2. \qquad \square \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

Note that this rate estimate is of limited use, since it involves the estimated parameter value α^*, which is not known a priori. We expect that a deeper convergence rate analysis is more involved for the GCV-functional. One reason for this is that the GCV-functional is designed to estimate the residual $\|A_n x_{\alpha,\delta,n} - y_n\|$ rather than the expression that we actually want to minimize, $\|x_{\alpha,\delta,n} - x^\dagger\|$. A further analysis would require to bound α^* in terms of the optimal α, but quite probably this needs more requirements than just noise and regularity conditions. An extensive analysis—even proving optimal order estimates—for the GCV-functional was done by M. Lukas; see, e.g., [12, 13, 14]. This was shown using quite specific bounds on the rates of decay of the noise components, the components of the exact solution and the decay of the singular values. Moreover, experiments comparing parameter choice rules indicated that the GCV-functional usually performs worse than the other rules.

5. Noise condition and random noise. We now study the discrete noise conditions (4.1), (4.2) in more detail for the case of random noise. A highly relevant noise model is white noise for which we derive probabilistic estimates for the noise conditions. If additionally $x_n^\dagger$ satisfies one of the assumptions of Theorem 4.8, this leads to the result that the bounds (4.11), (4.12), or (4.13) hold with the same probability as the noise condition.

We assume that all Fourier components of the noise are independent and normally distributed with mean zero and unknown variance σ,

$$(y_\delta - y, v_i) = e_i \sim N(0, \sigma^2), \qquad \text{independent}, \qquad i = 1, \ldots N. \tag{5.1}$$

LEMMA 5.1. *Let the white noise model* (5.1) *hold. For any* $\xi \in [\lambda_{min}, \lambda_{max}]$ *we define*

$$z(\xi) = \min\{i \in \mathbb{N} \,|\, \lambda_i \le \xi\}. \tag{5.2}$$

Suppose that for some $k, m \in \mathbb{N}$ *with* $m > 2k$ *there exists a constant* $\tilde{C}_m$ *such that for all* $\xi \in I_n$,

$$\xi^2 \sum_{i=1}^{z(\xi)-1} \frac{1}{\lambda_i} \le \tilde{C}_m \sum_{s=1}^{\lfloor \frac{N+1-z(\xi)}{m} \rfloor} \lambda_{N-m(s-1)}. \tag{5.3}$$

Then (4.1) *is satisfied with a constant* C_{ncd} *with probability* p,

$$p \ge 1 - \left(\prod_{i=1}^{k} \frac{2i-1}{m-2i} \right) \left(\frac{\tilde{C}_m}{C_{ncd}} \right)^k. \tag{5.4}$$

Moreover, if for some $k \in \mathbb{N}$ *there exists a constant* $\tilde{C}_k$ *such that for all* $\xi \in I_n$,

$$\xi \sum_{1=1}^{z(\xi)-1} \frac{1}{\lambda_i} \le \tilde{C}_k (N - z(\xi) + 1 - 2k), \tag{5.5}$$

then (4.2) *is satisfied with a constant* C_{ncd} *with probability* p,

$$p \ge 1 - \prod_{i=1}^{k} (2i-1) \left(\frac{\tilde{C}_k}{C_{ncd}} \right)^k. \tag{5.6}$$

Proof. The probability that (4.1) is satisfied with the constant C_{ncd} can be estimated as follows

$$\mathbb{P}((4.1)) = 1 - \mathbb{P}\left(\xi^2 \frac{\sum_{\lambda_i > \xi} \frac{1}{\lambda_i} e_i^2}{\sum_{\lambda_i \le \xi} \lambda_i e_i^2} \ge C_{ncd}, \text{ for some } \xi \in I_n \right).$$

ETNA
Kent State University
http://etna.math.kent.edu

Let $\xi \in I_n$ be fixed, $k \in \mathbb{N}$ be arbitrary. Then by Markov's inequality and independence,

$$\mathbb{P}\left(\xi^2 \frac{\sum_{\lambda_i>\xi} \frac{1}{\lambda_i} e_i^2}{\sum_{\lambda_i\leq\xi} \lambda_i e_i^2} \geq C_{ncd}\right) = \mathbb{P}\left(\left(\frac{\sum_{\lambda_i>\xi} \frac{1}{\lambda_i} e_i^2}{\sum_{\lambda_i\leq\xi} \lambda_i e_i^2}\right)^k \geq \frac{C_{ncd}^k}{\xi^{2k}}\right)$$
$$\leq \frac{\xi^{2k}}{C_{ncd}^k}\mathbb{E}\left(\left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}e_i^2\right)^k\right)\mathbb{E}\left(\left(\frac{1}{\sum_{\lambda_i\leq\xi}\lambda_i e_i^2}\right)^k\right).$$

By the convexity of the power function we obtain that

$$\mathbb{E}\left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}e_i^2\right)^k = \left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}\right)^k \mathbb{E}\left(\sum_{\lambda_i>\xi}\frac{\frac{1}{\lambda_i}}{\sum_{\lambda_j>\xi}\frac{1}{\lambda_j}}e_i^2\right)^k$$
$$\leq \left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}\right)^k \sum_{\lambda_i>\xi}\frac{\frac{1}{\lambda_i}}{\sum_{\lambda_j>\xi}\frac{1}{\lambda_j}}\mathbb{E}e_i^{2k} = \left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}\right)^k \sigma^{2k}2^k\frac{\Gamma(k+\frac{1}{2})}{\Gamma(\frac{1}{2})}$$
$$= \left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}\right)^k \sigma^{2k}\frac{(2k)!}{2^k k!} = \left(\sum_{\lambda_i>\xi}\frac{1}{\lambda_i}\right)^k \sigma^{2k}\Pi_{i=1}^k(2i-1).$$

Now to the second factor. Since the λ_i are ordered monotonically decreasing, we can estimate for arbitrary but fixed $m \in \mathbb{N}$, $m > 2k$, with $z = z(\xi)$ as in (5.2),

$$\sum_{\lambda_i\leq\xi}\lambda_i e_i^2 = \sum_{i=z}^N \lambda_i e_i^2 = \sum_{i=1}^{N+1-z}\lambda_{N+1-i}e_{N+1-i}^2$$
$$\geq \sum_{l=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\sum_{k=1}^m \lambda_{N+1-(m(l-1)+k)}e_{N+1-(m(l-1)+k)}^2$$
$$\geq \sum_{l=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\lambda_{N+1-(m(l-1)+1)}\sum_{k=1}^m e_{N+1-(m(l-1)+k)}^2$$
$$= \sum_{s=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\lambda_{N-m(s-1)}\left(\sum_{l=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\tau_l\sum_{k=1}^m e_{N+1-(m(l-1)+k)}^2\right),$$

where we set

$$\tau_l = \frac{\lambda_{N-m(l-1)}}{\sum_{s=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\lambda_{N-m(s-1)}}.$$

Obviously it hold that $\sum_{l=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\tau_l = 1$ and hence we can use the convexity of the function $\frac{1}{x^k}$ to find

$$\mathbb{E}\left(\left(\frac{1}{\sum_{\lambda_i\le\xi}\lambda_i e_i^2}\right)^k\right)$$
$$\le \frac{1}{\left(\sum_{s=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\lambda_{N-m(s-1)}\right)^k}\mathbb{E}\left(\frac{1}{\sum_{l=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\tau_l\sum_{k=1}^m e_{N+1-(m(l-1)+k)}^2}\right)^k$$
$$\le \frac{1}{\left(\sum_{s=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\lambda_{N-m(s-1)}\right)^k}\sum_{l=1}^{\lfloor\frac{N+1-z}{m}\rfloor}\tau_l\mathbb{E}\left(\frac{1}{\sum_{k=1}^m e_{N+1-(m(l-1)+k)}^2}\right)^k.$$

The last factor can be calculated using the noise model (5.1),

$$\mathbb{E}\left(\frac{1}{\sum_{k=1}^m e_{N+1-(m(l-1)+k)}^2}\right)^k = \frac{1}{\sigma^{2k}}\frac{1}{2^k}\frac{\Gamma(\frac{m}{2}-k)}{\Gamma(\frac{m}{2})} = \frac{1}{\sigma^{2k}}\frac{1}{\Pi_{j=1}^k(m-2j)}.$$

Combining the estimates yields the result for (4.1). We can follow the main steps also for the case of (4.2). The only difference is that we have to estimate

$$\mathbb{E}\left(\left(\frac{1}{\sum_{\lambda_i\le\xi}e_i^2}\right)^k\right) = \frac{1}{\sigma^{2k}}\frac{1}{2^k}\frac{\Gamma(\frac{N-z+1}{2}-k)}{\Gamma(\frac{N-z+1}{2})} = \frac{1}{\sigma^{2k}}\frac{1}{\Pi_{j=1}^k(N-z+1-2j)}$$
$$\le \frac{1}{\sigma^{2k}}(N-z+1-2k)^{-k},$$

where we have to restrict ourselves to $2k < N-z+1$. □

Let us remark that the coefficients m, k are tuning parameters, to make the probability as high as possible. In particular, if $\tilde{C}_m$ and $\tilde{C}_k$ are known, one can try to set $m = 2k+1$ and maximize the probability estimates over k, if the simplest choice $k = 1, m = 3$ does not yield appropriate estimates of the probability. Moreover, we may also vary the constant C_{ncd} to conclude that (4.11)–(4.13) hold with a constant C with certain probability. The results of Lemma 5.1 in combination with Theorem 4.8 are consistent with what one would expect: the bounds on the right-hand side of (5.4), (5.6) are monotonically increasing with C_{ncd} and so are the constants C in (4.11)–(4.13). Thus a good error bound with small C holds with a lower probability than a bad error bound with large C. Note that the right-hand sides in (5.4), (5.6) may become negative, in which case the lemma is vacuous.

5.1. Case studies. We can further investigate the constants in (5.3) and (5.5) when a certain decay of the singular values and λ_i is assumed. In particular, we want to investigate about the interval I_n, where these inequalities hold with moderate and discretization independent constants. It turns out that in many cases an interval of the form

$$\xi \in I_n = [\lambda_z, \lambda_1]$$

can be taken, where $1 \le z < N$ and N is the total number of singular values. Below we derive appropriate estimates for the index z leading to discretization independent bounds.

We assume that

$$\sigma_i^2 = \lambda_i = \phi(i),$$

ETNA
Kent State University
http://etna.math.kent.edu

with a monotonically decreasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$. In this case we can estimate the sums in (5.3) and (5.5) by integrals:

$$\sum_{1=1}^{z(\xi)-1} \frac{1}{\lambda_i} \leq \int_1^{z(\xi)-1} \frac{1}{\phi(t)}\,dt + \frac{1}{\phi(z(\xi)-1)} \leq \int_1^{z(\xi)} \frac{1}{\phi(t)}\,dt, \tag{5.7}$$

$$\sum_{s=1}^{\lfloor \frac{N+1-z(\xi)}{m} \rfloor} \lambda_{N-m(s-1)} \geq \int_1^{\lfloor \frac{N+1-z(\xi)}{m} \rfloor + 1} \phi(N-m(s-1))\,ds$$

$$= \frac{1}{m}\int_{N-m\lfloor \frac{N+1-z(\xi)}{m} \rfloor}^{N} \phi(\tau)\,d\tau \geq \frac{1}{m}\int_{z(\xi)-1+m}^{N} \phi(\tau)\,d\tau, \tag{5.8}$$

$$\xi^2 \leq \phi(z-1)^2. \tag{5.9}$$

Case 1. Mildly ill-posed problems.
Assume a typical example of a mildly ill-posed problem, where the rate of the decay of the singular values is polynomial, i.e., $\phi(t) = \frac{1}{t^p}$ with $p > 1$. Then the constant in (5.3) can be estimated by bounding the ratio (using (5.7)–(5.9))

$$\begin{aligned} \frac{\xi^2 \sum_{1=1}^{z(\xi)-1} \frac{1}{\lambda_i}}{\sum_{s=1}^{\lfloor \frac{N+1-z(\xi)}{m} \rfloor} \lambda_{N-m(s-1)}} &\leq m\frac{p-1}{p+1} \frac{z^{p+1}-1}{(z-1)^{2p}\left(\frac{1}{(z-1+m)^{p-1}} - \frac{1}{N^{p-1}}\right)} \\ &= m\frac{p-1}{p+1}\frac{z^{p+1}-1}{(z-1)^{p+1}}\left(\frac{z-1+m}{z-1}\right)^{p-1} \frac{1}{1-\left(\frac{z-1+m}{N}\right)^{p-1}} \\ &\leq m\frac{p-1}{p+1}\left(1+\frac{1}{z-1}\right)^{p+1}\left(1+\frac{m}{z-1}\right)^{p-1}\frac{1}{1-\left(\frac{z-1+m}{N}\right)^{p-1}}. \end{aligned} \tag{5.10}$$

In the same way we obtain for (5.5),

$$\frac{1}{p+1}\frac{1}{(z-1)^p}\frac{(z+1)^{p+1}-1)}{N-z-2k} \leq \frac{1}{p+1}\left(\frac{z+1}{z-1}\right)^p \frac{1}{\frac{N}{z}-1-\frac{2k}{z}}.$$

Thus, if $\theta < 1$ and $z_0 > 1$, and if ξ is such that

$$1 + \frac{m}{z_0^{\frac{1}{p+1}} - 1} \leq z(\xi) \leq 1 + \theta N - m, \tag{5.11}$$

then we find that with

$$\tilde{C}_m = z_0^{2p} m \frac{p-1}{p+1}\frac{1}{1-\theta}$$

that (5.3) is satisfied, and if

$$1 + \frac{1}{2(z_0^{\frac{1}{p}} - 1)} \leq z(\xi) \leq (\theta+1)(N-2k),$$

then (5.5) holds with a constant

$$\tilde{C}_m \leq \frac{1}{p+1} z_0^p \frac{1}{\theta}.$$

ETNA
Kent State University
http://etna.math.kent.edu

Case 2. Exponentially ill-posed problems.
We assume that $\phi(t) = e^{-\gamma t}$ for some $\gamma > 0$. Estimating the ratio as in (5.10) by integrals yields

$$m(e^{-\beta(z-1)})^2 \frac{e^{\beta z} - e^{\beta}}{e^{-\beta(z-1+m)} - e^{-\beta N}} \leq m \frac{e^{\beta}}{e^{-\beta m} - e^{-\beta(N-(z-1)}} \leq m \frac{e^{\beta(m+1)}}{1 - e^{-\beta(N-(z-1)+m}} \; .$$

Hence, if $z(\xi)$ is such that

$$z(\xi) \leq N + 1 - m - \gamma,$$

then (5.3) is satisfied with an N-independent constant

$$\tilde{C}_m = m \frac{e^{\beta(m+1)}}{1 - e^{-\beta\gamma}} \; .$$

The same holds for (5.5), since (5.3) implies (5.5).

Case 3. The backward heat equation.
Here we assume that the decay rate of the singular values is like that of the backward heat equation: $\phi(t) = e^{-\beta t^2}$, $\beta > 0$.

Let us first consider (5.5). The ratio (5.10) can be bounded by

$$\frac{1}{N + 1 - z - 2k} e^{-\beta(z-1)^2} \int_0^{z-1} e^{-\beta y^2} dy + e^{-\beta(z-1)} \leq \frac{1 + \frac{1}{\beta}}{N + 1 - z - 2k} \; .$$

This follows from estimates for the so-called Dawson integral,

$$e^{-x^2} \int_0^x e^{y^2} dy \leq 1,$$

(which also can be bounded by $\frac{1}{x}$). Hence, if

$$z(\xi) \leq N + 1 - 2k - \theta,$$

then (5.5) holds with constant

$$\tilde{C}_k = \frac{1 + \frac{1}{\beta}}{\theta} \; .$$

The situation for (5.3) is different in this case. The reason is that we were not able to find lower bounds for the denominator of (5.10),

$$e^{\beta(z-1)^2} \int_{z-1+m}^{N} e^{-\beta\tau^2} d\tau$$

by an N-independent constant. In fact, for $m > 0$ this value is bounded from above by an exponentially decaying term $e^{-\beta z}$ as $z \to \infty$, which cannot be compensated by a similar factor in the numerator. Hence, there is little hope that (5.3) holds for the backward heat equation. We can at best take an interval where ξ is within the first few eigenvalues $[\lambda_3, \lambda_1]$ (note the extremely fast decay of the singular value), which by construction yields a (maybe bad) but N-independent constant.

ETNA
Kent State University
http://etna.math.kent.edu

6. Discussion. We have established a convergence theory for certain noise level-free parameter choice rules in the discrete case. Here we focused on the quasioptimality principle and the Hanke-Raus rules. The generalized cross validation does not fit well into our analysis, and we believe that it is in practice inferior to the other rules in the case of ill-conditioned problems.

Comparing the results, we observed that the quasioptimality principle yields the best estimates (of oracle type), while the HR_1 rule has the bound $\frac{\delta}{\sqrt{\alpha}}$ instead of the propagated error term in Theorem 4.8. The HR_∞ rule is even worse. Although these factors are all of the same order, this gives a hint that the quasi-optimality rule has a lower error in many situations. This can also be seen from the experimental results in [16], where the HR_1 rule performs well but its error is usually larger by a factor (less than 10) than that of the quasioptimality rule. However, it is important to notice that the Hanke-Raus rules require a weaker noise condition than the quasioptimality rule.

Our analysis sheds light on the question, how the parameter choice rules should be implemented. The main problem is to specify what we consider an appropriate minimum, i.e., how to select the interval I_n in (2.4) to avoid the choice $\alpha^* = 0$. This issue might look like a minor one, but we believe it is the reason, that in case studies [4, 8, 16] the experimentally observed convergence results for noise level-free parameter choice rules are not always conclusive.

Let us discus several proposed rules-of-thumb for the practical implementation. For instance, Hansen and Hanke [8] proposed to use $I_n = [\alpha_{min}, \lambda_{max}]$, where α_{min} is a point of a "peak", i.e., a local maximum of the parameter choice functional. Such a maximum will usually appear only at $\alpha_{min} > \lambda_{min}$ and hence it automatically rules out the choice $\alpha^* = 0$. The problem here is that a maximum at the lower eigenvalues is not a very stable quantity because it might disappear when the structure of the noise at the lower eigenvalues is perturbed only a tiny bit. Variants of this idea are to quantify the meaning of an appropriate maximum and only consider those α_{min}, where the maximum has a relative large value, controlled by some factors (see the work of Palm, Hämarik and Raus [6, 7, 16]; compare also the climbing approach [16]). In general this is not a bad idea, since the appearance of a peak close to λ_{min} is a strong hint that a noise condition is satisfied. However, even if no such peak appears, heuristic parameter choice methods can be successful. The "to the right of the first peak"-rule does not indicate what to do then.

What we propose instead is to look for α^* in a fixed interval I_n, thus the optimum α^* does not need to be an interior minimum but can also be at the end points of I_n. Of crucial importance is the choice of this interval, since it has to be related to the noise-structure and the decay of the singular values. One suggestion in the literature is to minimize over the interval $[\gamma\sigma_{min}^2, 1]$ with a chosen factor $\gamma > 1$. This has been proposed by Neubauer [15] and Palm [16]. Again, this is in many cases appropriate (for instance, it agrees with our choice in the case of mildly ill-posed problems) but may fail in practically relevant situations. Instead, the analysis in Section 5.1 suggests the following approach: choose the interval I_n as $[\lambda_{n_{max}}, \lambda_{n_{min}}]$ for some integers n_{min}, n_{max}, $1 \leq n_{min} < n_{max} < N$ that can be derived from the lower and upper bounds $n_{min} \leq z(\xi) \leq n_{max}$ in Section 5.1. This means that the interval is constructed via the index set $i = 1, \ldots N$ of λ_i. Doing so (or similarly) is a quite reliable choice that works in many cases. However, as the previous analysis indicates, the quasioptimality principle fails even with this choice for problems of the type of the backward heat equation (when the singular values decay as e^{-i^2}). Here, only the Hanke-Raus rules work well. This observation is underpinned by empirical facts: the study in [16] showed that the quasioptimality principle has a very large error for the backward heat equation problem. A further source of failure could be that the lower bound for z in (5.11) is not independent of p, i.e., for mildly ill-posed problem with large p we have to define I_n by cutting out some of the *larger singular values* (depending on p) additionally to the smaller ones: $n_{min} = n_{min}(p) > 1$. It is questionable if this is of much importance in practice, since such a situation is only relevant when the optimal regularization parameter is close to the largest singular values. This only happens when the noise itself is of the order of the first singular values.

Let us mention that the analysis of Section 5 is also possible in the case of colored noise, $e_i \sim N(0, \sigma_i)$, when the decay of σ_i is known. The only difference is that the sums in (5.3), (5.5) have to be scaled appropriately by $\frac{\lambda_i}{\sigma_i}$ or $\lambda_i \sigma_i$. By a unitary transformation, the case of correlated noise $(e_i)_i \sim N(0, \Gamma)$ with known covariance matrix Γ can be handled as well.

We did not discuss the regularity condition too much, since it does not seem to be too important for choosing I_n. The comments in [10] about this issue are also adequate in the discrete case.

REFERENCES

[1] M. A. ARIÑO AND B. MUCKENHOUPT, *Maximal functions on classical Lorentz spaces and Hardy's inequality with weights for nonincreasing functions*, Trans. Amer. Math. Soc., 320 (1990), pp. 727–735.

[2] F. BAUER AND S. KINDERMANN, *The quasi-optimality criterion for classical inverse problems*, Inverse Problems, 24 (2008), 035002 (20 pages).

[3] ———, *Recent results on the quasi-optimality principle*, J. Inverse Ill-posed Probl., 17 (2009), pp. 1129–1142.

[4] F. BAUER AND M. A. LUKAS, *Comparing parameter choice methods for regularization of ill-posed problems*, Math. Comput. Simulation, 81 (2011), pp. 1795–1841.

[5] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[6] U. HÄMARIK, R. PALM, AND T. RAUS, *On minimization strategies for choice of the regularization parameter in ill-posed problems*, Numer. Funct. Anal. Optim., 30 (2009), pp. 924–950.

[7] ———, *Extrapolation of Tikhonov regularization method*, Math. Model. Anal., 15 (2010), pp. 55–68.

[8] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Survey Math. Indust., 3 (1993), pp. 253–315.

[9] M. HANKE AND T. RAUS, *A general heuristic for choosing the regularization parameter in ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 956–972.

[10] S. KINDERMANN, *Convergence Analysis for minimization based noise level-free parameter choice rules for linear ill-posed problems.* Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257. http://etna.mcs.kent.edu/vol.38.2011/pp233-257.dir/pp233-257.html

[11] S. KINDERMANN AND A. NEUBAUER, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging, 2 (2008), pp. 291–299.

[12] M. A. LUKAS, *Asymptotic optimality of generalized cross-validation for choosing the regularization parameter*, Numer. Math., 66 (1993), pp. 41–66.

[13] ———, *Comparisons of parameter choice methods for regularization with discrete noisy data*, Inverse Problems, 14 (1998), pp. 161–184.

[14] ———, *Robust generalized cross-validation for choosing the regularization parameter*, Inverse Problems, 22 (2006), pp. 1883–1902.

[15] A. NEUBAUER, *The convergence of a new heuristic parameter selection criterion for general regularization methods*, Inverse Problems, 24 (2008). 055005 (10 pages).

[16] R. PALM, *Numerical Comparison of Regularization Algorithms for Solving Ill-Posed Problems*, Ph.D. Thesis, Institute of Computer Science, University of Tartu, Tartu, Estonia, 2010.

[17] T. SEIDMAN, *Nonconvergence results for the application of least-squares estimation to ill-posed problems*, J. Optim. Theory Appl., 30 (1980), pp. 535–547.

[18] V. D. STEPANOV, *The weighted Hardy's inequality for nonincreasing functions*, Trans. Amer. Math. Soc., 338 (1993), pp. 173–186.

[19] A. TIKHONOV AND V. GLASKO, *The approximate solution of Fredholm integral equations of the first kind*, Zh. Vychisl. Mat. i Mat. Fiz., 4 (1964), pp. 564–571.

[20] ———, *Use of the regularization method in non-linear problems*, U.S.S.R. Comput. Math. Math. Phys., 5 (1965), pp. 93–107.

[21] G. WAHBA, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.