ISSN 1066-5307, Mathematical Methods of Statistics, 2014, Vol. 23, No. 3, pp. 176–200. © Allerton Press, Inc., 2014.

General Regularization Schemes for Signal Detection in Inverse Problems

C. Marteau[1*] and P. Mathé[2]**

[1]Inst. Math. de Toulouse, INSA de Toulouse, Univ. de Toulouse, France

[2]Weierstrass Inst. for Appl. Anal. and Stoch., Berlin, Germany

Received April 14, 2014; in final form, September 10, 2014

Abstract—The authors discuss how general regularization schemes, in particular, linear regularization schemes and projection schemes, can be used to design tests for signal detection in statistical inverse problems. It is shown that such tests can attain the minimax separation rates when the regularization parameter is chosen appropriately. It is also shown how to modify these tests in order to obtain a test which adapts (up to a log log factor) to the unknown smoothness in the alternative. Moreover, the authors discuss how the so-called *direct* and *indirect* tests are related in terms of interpolation properties.

Keywords: inverse problems, inverse test, linear regularization, projection scheme.

2000 Mathematics Subject Classification: 62G05, 62K20.

DOI: 10.3103/S1066530714030028

Dedicated to the memory of Yuri I. Ingster who passed away too early at the age of 64.

1. INTRODUCTION AND MOTIVATION

Statistical inverse problems have been intensively studied over the last years. Mainly, estimation of indirectly observed signals was considered. On the other hand, there are only a few studies concerned with signal detection, which is a problem of statistical testing. This is the core of the present paper.

1.1. The Model

Precisely, we consider a statistical problem in Hilbert space, where we are given two separable and real Hilbert spaces H and K, with norms denoted by $\|\cdot\|_H$, $\|\cdot\|_K$, respectively, along with a (compact) linear operator $T\colon H \to K$. Given the (unknown) element $f \in H$ we observe

$$Y = Tf + \sigma\xi, \tag{1.1}$$

where ξ is a Gaussian white noise, and σ is a positive noise level. A large amount of attention has been payed to the estimation issue, where one wants to estimate the function f of interest, and to control the associated error. We refer for instance to [9] for a review of existing methods in a deterministic setting (ξ is a deterministic error satisfying $\|\xi\|_K \leq 1$). In the statistical framework, the noise ξ is not assumed to be bounded. In this case, there is a slight abuse of notation in using (1.1). We assume in fact that for all $g \in K$, we can observe

$$\langle Y, g\rangle = \langle T, f\rangle g + \sigma\langle \xi, g\rangle, \tag{1.2}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product in K. Details will be given in Section 2. In this context, we mention [3] or [7] among others for a review of existing methodologies and related rates of convergence for estimation under Gaussian white noise.

*E-mail: clement.marteau@math.univ-toulouse.fr

**E-mail: peter.mathe@wias-berlin.de

1.2. Setup of (Nonparametric) Statistical Tests

Our aim is to test the null hypothesis that the (underlying true) signal f corresponds to a given signal f_0 against a nonparametric alternative. More formally, we test

$$H_0\colon f = f_0, \quad \text{against } H_{1,\rho}\colon f - f_0 \in \mathcal{E}, \quad \|f - f_0\|_H \geq \rho, \tag{1.3}$$

where $\mathcal{E}$ is a subset of H, and $\rho > 0$ a given radius. The subset $\mathcal{E}$ can be understood as a smoothness constraint on the remainder $f - f_0$, while the quantity ρ measures the amount of signal, different from f_0, available in the observation. The setting (1.3) is known as a goodness-of-fit or a signal detection (when $f_0 = 0$) testing problem. Clearly, since f_0 and hence Tf_0 are given, we can confine the analysis to testing whether $f = 0$ (no signal) against the alternative $H_{1,\rho}\colon f \in \mathcal{E}$, $\|f\|_H \geq \rho$, and we discuss this simplified model from now on.

In the following, we will deal with level-α tests, i.e., measurable functions of the data with values in $\{0, 1\}$. By convention, we reject H_0 if the test is equal to 1 and do not reject this hypothesis, otherwise. We are interested in the optimal value of ρ (see (1.3)) for which a prescribed level for the error of the second kind can be attained. More formally, given a fixed value of $\beta \in (0, 1)$ and a level-α test Φ_α, we are interested in the radius $\rho(\Phi_\alpha, \beta, \mathcal{E})$ defined as

$$\rho(\Phi_\alpha, \beta, \mathcal{E}) = \inf\Big\{\rho \in \mathbb{R}^+\colon \sup_{f\in\mathcal{E},\ \|f\|_H > \rho} P_f(\Phi_\alpha = 0) \leq \beta\Big\}.$$

From this, the minimax separation radius $\rho(\alpha, \beta, \mathcal{E})$ can be defined as the smallest radius over all possible testing procedures, i.e.,

$$\rho(\alpha, \beta, \mathcal{E}) = \arg\min_{\Phi_\alpha} \rho(\Phi_\alpha, \beta, \mathcal{E}),$$

and the minimum is over all level-α tests Φ_α. We stress that this minimax separation radius will depend on the noise level σ, and on spectral properties, both of the operator T which governs the equation (1.1), and of the class $\mathcal{E}$ describing the smoothness of the alternative.

1.3. State of the Art and Objective of the Study

In the direct case, i.e., when $T = I$, this problem has been widely investigated. We mention for instance seminal investigations proposed in [12–14]. We refer also to [1] where a non-asymptotic approach is proposed. Concerning testing in inverse problems, there exists, up to our knowledge, only a few references, as, e.g., [15] and [18]. In these contributions, a preliminary estimator $\hat{f}$ for the underlying signal f is used. This estimator is based on a spectral cut-off scheme in [18], or on a refined version using Pinsker's filter in [15]. All these approaches are based on a truncation of the singular value decomposition (svd).

By using such preliminary estimator the separation radius was upper and lower bounded under various smoothness assumptions (see for instance [15] or [18]). However, the restriction to truncated svd narrows the applicability of the test procedures, since a singular value decomposition is often hardly available, for instance, when considering partial differential equations on domains with noisy boundary data.

In order to motivate our subsequent discussion, we briefly recall the truncated svd approach. Suppose that the operator T has a singular value decomposition $(s_j, u_j, v_j)_{j\in\mathbb{N}^*}$, where $\mathbb{N}^\star := \mathbb{N}\backslash\{0\}$, $(s_j^2)_{j\in\mathbb{N}^*}$ denotes the eigenvalues of T^*T and $(u_j)_{j\in\mathbb{N}^*}$ (resp. $(v_j)_{j\in\mathbb{N}^*}$) an orthonormal set in H (resp. K). Then, for all $x \in H$, $Tx = \sum_{j=1}^{\infty} s_j \langle x, u_j\rangle_H\ v_j$. In this context, the truncated svd consists in the choice of an integer D and letting

$$\hat{f} = \hat{f}_D := \sum_{j=1}^{D} \frac{1}{s_j}\langle Y, v_j\rangle u_j, \qquad Y \in K. \tag{1.4}$$

This reconstruction may be viewed from two perspectives. First, it uses discrete data $\langle Y, v_j\rangle$, $j = 1, \ldots, D$, and it represents the solution with respect to the finite system $u_1, \ldots, u_D$. It thus corresponds to a specific instance of *projection schemes* (parametrized through the dimensionality D), well studied in regularization theory. However, the choice of D may also be obtained from truncating the sequence of singular numbers as $D := \#\{j,\ s_j^2 > \tau\}$ with a truncation parameter $0 < \tau \le \|T^*T\|$, $\|\cdot\|$ denoting the operator norm. From this perspective, truncated svd may be regarded as a special instance of *linear regularization*. A prominent example for the latter is Tikhonov regularization, in which case we let $\hat{f} = \hat{f}_\tau = (\tau I + T^*T)^{-1}T^*Y$, $Y \in K$, but several alternative regularization schemes are available in the literature (see, e.g., [9]). Thus the question arises which requirements are to be put on projection schemes and/or linear regularization in order to be used for inverse testing. We will highlight these requirements in our subsequent discussion. For the time being we mention that attention must be paid to the capability of the chosen reconstruction scheme to use the inherent solution smoothness in an (order) optimal way. Within the recent theory of inverse problems the smoothness, which is inherent in the class $\mathcal{E}$, is measured *relative* to the operator T. By doing so, a unified treatment of moderately, severely and mildly ill-posed problems is possible. We take this paradigm here and consider the classes $\mathcal{E}$ as *source sets*, see details in § 4. We shall establish results and conditions such that these alternative testing procedures match the previous minimax bounds.

There is an interesting relation between testing based on the estimation of f (inverse test), and test based on the estimation of Tf (direct test). Such discussion can already be found in [17]. However, here we highlight that the relation between both problems can be seen as a result of *interpolation* between smoothness spaces, the one which describes the signal f and the one which characterizes the smoothness of Tf.

Finally, we shall establish an adaptive test, which is based on a given finite family $\mathcal{R}$ of non-adaptive tests. It will be shown that this adaptive test does no longer use any a priori smoothness index. At the same time it is, up to a log log factor, as good (in terms of the error of the second kind) as the best test among the whole family $\mathcal{R}$.

1.4. Outline

We first consider general linear reconstruction mappings $\hat{f}(Y) = R(Y)$, $Y \in K$, in terms of an operator R (Sections 3.1–3.2). In order to control the errors of the first and second kind, such reconstruction R needs to be a member of a parametrized family of mappings. Therefore, these families will be specified as linear regularization (in Section 4.1), or projection schemes (in Section 4.2), respectively. In each case, we derive the corresponding minimax separation radii.

Then we turn to discussing the relation between direct and inverse test problems in Section 5, and we describe an adaptive test and its properties in Section 6.

2. THE STATISTICAL INVERSE PROBLEM MODEL

2.1. Notation and Assumptions

First we will specify the assumption on the noise in (1.1).

Assumption A1 (Gaussian white noise). *The noise ξ is a weak random element in K, which has finite weak absolute second moments. Specifically, for all $g, g_1, g_2 \in K$, we have*

$$\langle \xi, g\rangle \sim \mathcal{N}(0, \|g\|_K^2) \quad (\text{and} \quad \mathbb{E}[\langle \xi, g_1\rangle\langle \xi, g_2\rangle] = \langle g_1, g_2\rangle).$$

Notice that the second property is a consequence of the first, because bilinear forms in Hilbert space are determined by their values at the diagonal.

We make the following assumptions for the operator T governing Eq. (1.1).

Assumption A2 (Singular value decomposition). *Let $(s_j, u_j, v_j)_{j\in\mathbb{N}^*}$ be the singular value decomposition of the (compact) operator T, where $(s_j^2)_{j\in\mathbb{N}^*}$ denotes the sequence of eigenvalues of T^*T, arranged in decreasing order, and $(u_j)_{j\in\mathbb{N}^*}$ (resp. $(v_j)_{j\in\mathbb{N}^*}$) are orthonormal systems in H (resp. K). In particular, we have*

$$Tf = \sum_{j=1}^{\infty} s_j \langle f, u_j\rangle_H v_j, \qquad f \in H.$$

We refer to [9] for more details regarding the construction of the singular value decomposition and related properties. In order that the estimators which are constructed below are well defined we impose the following assumption.

Assumption A3. *The operator T is a Hilbert–Schmidt operator, i.e.,*

$$\operatorname{tr}[T^*T] < +\infty.$$

As was already mentioned, we shall measure the smoothness relative to the operator T, and this is done as follows. Since the operator T is compact, so is the self-adjoint companion T^*T. The range of T^*T is a (dense) subset in H, and one may consider an element f smooth, if it is in the range of T^*T. To be more flexible, we shall do this as follows. To a bounded non-negative real function $\varphi\colon [0, \|T^*T\|] \to \mathbb{R}^+$ we assign the self-adjoint operator $\varphi(T^*T)$ using spectral calculus. The corresponding linear operator $\varphi(T^*T)$ is compact whenever $\varphi(t) \to 0$ as $t \to 0$. Therefore, we shall restrict considerations to functions with this property.

Assumption A4 (Source set). *For a continuous non-decreasing function φ with $\varphi(0) = 0$ (index function) we let*

$$\mathcal{E}_\varphi = \{h \in H,\ h = \varphi(T^*T)\omega,\ \text{for some } \|\omega\|_H \le 1\} \tag{2.1}$$

be a general source set.

Up to our knowledge, most of the contributions on testing theory proposed in the literature deal with smoothness constraints expressed through the svd of the operator T. More precisely, the set $\mathcal{E}$ is characterized by a constraint on the decay of the coefficients of f in the basis $(u_j)_j$. We refer to [18] or [15] for more details. Examples 4–5 relate Sobolev type balls to the present setup.

It was established in [20] that each element in H has some smoothness of the form (2.1), and hence the present approach is more general.

Under the above Assumption A1 on the noise, given any linear reconstruction operator $R\colon K \to H$ the element RY belongs to H almost surely, provided that R is a Hilbert–Schmidt operator (Sazonov's Theorem). When specifying the reconstruction R in Sections 4.1–4.2, we shall always make sure that this is the case.

Then the application of R to the data Y may be decomposed as

$$RY = RTf + \sigma R\xi = f_R + \sigma R\xi, \qquad f \in H, \tag{2.2}$$

where $f_R := RTf$ denotes the noiseless (deterministic part) of RY. Along with the reconstruction RY the following quantities will prove important. First, we can compute the bias–variance decomposition

$$\mathbb{E}\|RY\|_H^2 = \|RTf\|_H^2 + \sigma^2\mathbb{E}\|R\xi\|_H^2 = \|f_R\|_H^2 + S_R^2, \tag{2.3}$$

where we introduce the *variance* of the estimator RY as

$$S_R^2 := \sigma^2\mathbb{E}\|R\xi\|_H^2 = \sigma^2 \operatorname{tr}[R^*R], \tag{2.4}$$

which is finite if R is a Hilbert–Schmidt operator. In addition, the following *weak variance* will play a role:

$$v_R^2 := \sigma^2 \sup_{\|w\|_H \le 1} \mathbb{E}|\langle R\xi, w\rangle|^2 = \sigma^2\|R\|^2, \tag{2.5}$$

the latter norm denoting the operator norm of the mapping $R\colon K \to H$. Below, if R is clear from the context we sometimes abbreviate $S = S_R$ and $v = v_R$.

We will need more precise representation of the trace and norm as above in terms of the representation of the operator R. Suppose that R is given in terms of its singular value decomposition as

$$Rg = \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, g\rangle \phi_j, \qquad g \in K, \tag{2.6}$$

where we assume that both sequences ψ_j, ϕ_j, $j = 1, 2, \ldots$, are orthonormal bases in K and H, respectively. Moreover, the sequence λ_j, $j = 1, 2, \ldots$, is assumed nonnegative and arranged in non-increasing order. Then the following is well known.

Lemma 2.1. *Let R be as in* (2.6). *Then*

1. $\operatorname{tr}[R^*R] = \sum_{j=1}^{\infty} \lambda_j^2$, *and*

2. $\|R\|^2 = \sup_{j=1}^{\infty} \lambda_j^2$.

From this we can see that $v_R^2 \le S_R^2$.

2.2. The svd Paradigm

In the inverse problem literature, a large attention has been paid to algorithms based on the singular value decomposition (introduced in Assumption A2), and we already resorted to a brief discussion in § 1.3. Indeed, for all $j \in \mathbb{N}^*$, replacing g by v_j in (1.2), we get

$$\langle Y, v_j \rangle = s_j \langle f, u_j \rangle_H + \sigma \xi_j,$$

where the ξ_j are i.i.d. Gaussian random variables. This particular formulation of the model (1.1) is known as the *sequence space model*. We refer for instance to [7] for an extended review on estimation algorithms and related properties in such a setting. We specify the previously introduced strong and weak variances for the truncated svd estimator.

Example 1. Suppose that we approximate the inverse mapping of T, with svd given in Assumption A2, by the finite expansion from (1.4). Then the singular numbers λ_j of R obey $\lambda_j = s_j^{-1}$, $j = 1, \ldots, D$, and $\lambda_j = 0$, $j > D$, and hence its strong and weak variances are readily evaluated as $S_D^2 = \sum_{j=1}^{D} s_j^{-2}$ and $v_D^2 = s_D^{-2}$.

Statistical tests for the signal detection problems were previously constructed in [15, 18]. Those constructions use the above truncated singular value decomposition. The choice of the truncation level is based on the smoothness inherent in the class $\mathcal{E}$. In particular, the analysis of the spectral cut-off regularization scheme for testing in inverse problems provided in [18] revealed the importance of the quantity

$$\rho_D := \left(\sum_{j=1}^{D} \frac{1}{s_j^4} \right)^{1/4}, \qquad D = 1, 2, \ldots \tag{2.7}$$

We mention the non-asymptotic lower and upper bounds, extending the setup from [18] to the case of smoothness measured in terms of (general) source sets, given as

$$\begin{aligned} \rho^2(\alpha, \beta, \mathcal{E}_\varphi) &\ge \sup_D \min\{c_{\alpha,\beta}^2 \rho_D^2, \varphi^2(s_D^2)\}, \\ \rho^2(\alpha, \beta, \mathcal{E}_\varphi) &\le \inf_D \left(C_{\alpha,\beta}^2 \rho_D^2 + \varphi^2(s_D^2) \right). \end{aligned} \tag{2.8}$$

Again, the numbers s_j, $j = 1, 2, \ldots$ denote the sequence of singular numbers of the operator T, arranged in decreasing order, and the function φ describes smoothness, cf. Assumptions A2 and A4.

Recall the strong and weak variances S_D^2 and v_D^2 from Example 1. If $\rho_D^2 \asymp S_D v_D$, or more explicitly, if

$$\sum_{j=1}^{D} \frac{1}{s_j^4} \asymp \frac{1}{s_D^2} \sum_{j=1}^{D} \frac{1}{s_j^2},$$

then the minimax separation rate can be expressed in terms of the strong and weak variances. This concerns only the decay rate of the singular numbers s_j of the operator T, and this holds for regularly varying singular numbers, but this also holds true for $s_j \asymp \exp(-\gamma j)$, $j = 1, 2, \ldots$, thus covering severely ill-posed problems. Remark that unlike the terms involved in (2.7), the quantities S_D^2 and v_D^2 have a clear interpretation.

Thus, when using more general linear regularization R, it is important to determine whether its strong and weak variances S_R^2, v_R^2, cf. (2.4)–(2.5), allow for similar bounds. Precisely, for general linear regularization R we shall reduce the problem of bounding the separation radius to an optimization problem similar to Eq. (2.8), see Eq. (4.2). In order to do so, we need requirements for the regularization, and these will be discussed in Sections 4.1 and 4.2, respectively.

3. CONSTRUCTION AND CALIBRATION OF THE TEST

3.1. Construction of the Test

In the literature (see, e.g., [12–14], [24], [1], or [15]), several tests are based on an estimator of $\|f\|_H^2$ ($\|f - f_0\|_H^2$ in the general case). Then, the idea is to reject H_0 as soon as this estimator becomes too large with respect to a prescribed threshold.

In order to estimate $\|f\|_H^2$, where $f \in H$, from the observations Y, cf. (1.1), we shall use a general linear reconstruction operator $R\colon K \to H$. We see from (2.3) that the quantity $\|RY\|_H^2 - S_R^2$ is an unbiased estimator for the norm of $\|f_R\|_H^2$ with $f_R = R(Tf)$. If R is chosen appropriately, this term is an approximation of $\|f\|_H^2$, whose value is of first importance when considering the problem (1.3). Therefore, we shall use a threshold for $\|RY\|_H^2 - S_R^2$ to describe the test.

Let $\alpha \in (0, 1)$ be the prescribed level for the first kind error. We define the test $\Phi_{\alpha,R}$ as

$$\Phi_{\alpha,R} = \mathbf{1}_{\{\|RY\|_H^2 - S_R^2 > t_{R,\alpha}\}}, \tag{3.1}$$

where $t_{R,\alpha}$ denotes the $(1-\alpha)$-quantile of the variable $\|RY\|_H^2 - S_R^2$ under H_0. Due to the definition of the threshold $t_{R,\alpha}$, the test $\Phi_{\alpha,R}$ is a level-α test. Indeed

$$P_{H_0}(\Phi_{\alpha,R} = 1) = P_{H_0}\big(\|RY\|_H^2 - S_R^2 > t_{R,\alpha}\big) = \alpha.$$

We emphasize that under H_0 the distribution of $\|RY\|_H^2 - S_R^2 = \sigma^2(\|R\xi\|^2 - \mathrm{tr}\,[R^*R])$ depends only on the chosen reconstruction R. Hence the quantile can be determined, at least approximately.

3.2. Control of the Error of the Second Kind

Here, our aim is to control the error of the second kind by some prescribed level $\beta > 0$. To this end, we have to exhibit conditions on f for which the probability $P_f(\Phi_{\alpha,R} = 0)$ will be bounded by β. By construction of the above test this amounts to bounding

$$\begin{aligned} P_f(\Phi_{\alpha,R} = 0) &= P_f(\|RY\|_H^2 - S^2 \le t_{R,\alpha}) \\ &= P_f(\|RY\|_H^2 - \mathbb{E}\|RY\|_H^2 \le t_{R,\alpha} + S^2 - \mathbb{E}\|RY\|_H^2) \\ &= P_f(\|RY\|_H^2 - \mathbb{E}\|RY\|_H^2 \le t_{R,\alpha} - \|f_R\|_H^2), \end{aligned} \tag{3.2}$$

where the latter follows from (2.3). In this section, we will investigate the lowest possible value of $\|f_R\|_H^2$ for which this probability can be bounded by β.

Let $\beta \in (0, 1)$ be fixed. For all $f \in H$, we denote by $t_{R,\beta}(f)$ the β-quantile of the variable $\|RY\|_H^2 - S_R^2$. In other words

$$P_f\big(\|RY\|_H^2 - \mathbb{E}\|RY\|_H^2 \le t_{R,\beta}(f)\big) = \beta. \tag{3.3}$$

Then, we get from (3.2) and (3.3) that $P_f(\Phi_{\alpha,R} = 0)$ will be bounded by β as soon as

$$t_{R,\alpha} - \|f_R\|_H^2 \le t_{R,\beta}(f) \Leftrightarrow \|f_R\|_H^2 \ge t_{R,\alpha} - t_{R,\beta}(f). \tag{3.4}$$

In order to conclude this discussion, we need an upper bound on $t_{R,\alpha}$ and a lower bound on $t_{R,\beta}(f)$. To this end we recall (2.6), and we expand the element RY as

$$\begin{aligned} RY = \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, Y\rangle \phi_j &= \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, Tf\rangle \phi_j + \sigma \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, \xi\rangle \phi_j \\ &= \sum_{j=1}^{\infty} \theta_j \phi_j + \sum_{j=1}^{\infty} \sigma_j \varepsilon_j \phi_j, \end{aligned} \tag{3.5}$$

where $\theta_j := \lambda_j \langle \psi_j, Tf\rangle$, $\sigma_j := \sigma\lambda_j$ for all $j \in \mathbb{N}$, and the ε_j are i.i.d. standard Gaussian random variables. With this notation we let

$$\Xi := \sum_{j=1}^{\infty} \sigma_j^4 + 2\sum_{j=1}^{\infty} \sigma_j^2 \theta_j^2. \tag{3.6}$$

The proof of the following result is provided in Appendix A.

Lemma 3.1. *Let the reconstruction R be given as in* (2.6)*, and let $x_\gamma := \log(1/\gamma)$ for all $\gamma \in (0,1)$. Then*

$$t_{R,\alpha} \le 2\sqrt{2x_\alpha} S_R v_R + 2v_R^2 x_\alpha,$$

and, by using the quantity Ξ from (3.6)*, we have*

$$t_{R,\beta}(f) \ge -2\sqrt{\Xi x_\beta}.$$

We are now able to find a condition on $\|f_R\|_H^2$ in order to control the error of the second kind. We introduce the quantity

$$C^*_{\alpha,\beta} = (4\sqrt{x_\beta} + 4\sqrt{2x_\alpha}), \tag{3.7}$$

which is a function of α and β, only.

Proposition 3.1. *Let us consider the test $\Phi_{\alpha,R}$ as introduced in* (3.1)*, and let*

$$r^2(\Phi_{\alpha,R},\beta) := C^*_{\alpha,\beta} S v + (4x_\alpha + 8x_\beta)v^2. \tag{3.8}$$

Then

$$\sup_{f, \|f_R\|_H^2 \ge r^2(\Phi_{\alpha,R},\beta)} P_f(\Phi_{\alpha,R} = 0) \le \beta.$$

Proof. The equation (3.4) provides a condition for which $P_f(\Phi_\alpha = 0) \le \beta$. Using Lemma 3.1, we see that this condition is satisfied as soon as

$$\|f_R\|_H^2 \ge 2\sqrt{\Xi x_\beta} + 2\sqrt{2x_\alpha} S v + 2v^2 x_\alpha.$$

Now we bound

$$\Xi = \sigma^4 \sum_{j=1}^{+\infty} \lambda_j^4 + 2\sigma^2 \sum_{j=1}^{+\infty} \lambda_j^2 \times \lambda_j^2 \langle \psi_j, Tf\rangle^2 \le S^2 v^2 + 2v^2 \|f_R\|_H^2.$$

Using the inequality $ab \le a^2/2 + b^2/2$ for all $a, b \in \mathbb{R}$, we get

$$2\sqrt{\Xi x_\beta} \le 2Sv\sqrt{x_\beta} + 2\sqrt{2x_\beta}\|f\|_H v \le 2Sv\sqrt{x_\beta} + \frac{1}{2}\|f\|_H^2 + 4x_\beta v^2.$$

In particular, condition (3.4) will be satisfied as soon as

$$\frac{1}{2}\|f_R\|_H^2 \ge (2\sqrt{x_\beta} + 2\sqrt{2x_\alpha})Sv + v^2(2x_\alpha + 4x_\beta).$$

□

Remark 3.1. Note that the condition on $\|f_R\|_H^2$ is (as most of the results presented below) non-asymptotic, i.e., we do not require that $\sigma \to 0$. Using the property $v \le S$, we can obtain the simple bound

$$r^2(\Phi_{\alpha,R},\beta) \le C_{\alpha,\beta} S v, \quad \text{where } C_{\alpha,\beta} = 4\sqrt{x_\beta} + 4\sqrt{2x_\alpha} + 4x_\alpha + 8x_\beta. \tag{3.9}$$

In an asymptotic setting, the value of the constant $C_{\alpha,\beta}$ may sometimes be improved. In particular, the bound $v \le S$ is rather rough. In many cases, we will only deal with the constant $C^*_{\alpha,\beta}$, and we refer to Lemma 4.2 (see Section 4 below).

4. DETERMINING THE SEPARATION RADIUS UNDER SMOOTHNESS

As discussed in the previous section, the error of the second kind is controlled as soon as $\|f_R\|_H^2 \geq C_{\alpha,\beta} Sv$. Nevertheless, the alternative in (1.3) is expressed in terms of a lower bound on $\|f\|_H^2$. In this section, we use the smoothness of f in order to propose an upper bound for the separation radius.

Using the triangle inequality, we obtain

$$\|f_R\|_H \geq \|f\|_H - \|f - f_R\|_H.$$

Hence $\|f_R\|_H^2 \geq r^2(\Phi_{\alpha,R}, \beta)$ as soon as

$$\begin{aligned}\|f\|_H - \|f - f_R\|_H \geq r(\Phi_{\alpha,R}, \beta) &\Leftrightarrow \|f\|_H^2 \geq \left(r(\Phi_{\alpha,R}, \beta) + \|f - f_R\|_H\right)^2, \\ &\Leftarrow \|f\|_H^2 \geq 2r^2(\Phi_{\alpha,R}, \beta) + 2\|f - f_R\|_H^2,\end{aligned}$$

In other words, we get from Proposition 3.1 that

$$\sup_{f, \|f\|_H^2 \geq 2r^2(\Phi_{\alpha,R},\beta) + 2\|f - f_R\|_H^2} P_f(\Phi_{\alpha,R} = 0) \leq \beta. \tag{4.1}$$

Hence we need to make the lower bound on $\|f\|_H$ as small as possible. We aim at finding sharp upper bounds for

$$\inf_{R \in \mathcal{R}} \left(r^2(\Phi_{\alpha,R}, \beta) + \|f - f_R\|_H^2\right), \tag{4.2}$$

where the reconstructions R belong to certain families $\mathcal{R}$.

4.1. Linear Regularization

We recall the notion of linear regularization, see, e.g., [11, Definition 2.2]. Such approaches are rather popular for estimation purpose. In this section, we describe how these schemes can be tuned in order to obtain suitable tests.

Definition 1 (Linear regularization)**.** A family of functions

$$g_\tau \colon (0, \|T^*T\|] \mapsto \mathbb{R}, \quad 0 < \tau \leq \|T^*T\|,$$

is called regularization if they are piece-wise continuous in τ and the following properties hold:

(1) For each $0 < t \leq \|T^*T\|$ we have $|r_\tau(t)| \to 0$ as $\tau \to 0$;

(2) There is a constant γ_1 such that $\sup_{0 \leq t \leq \|T^*T\|} |r_\tau(t)| \leq \gamma_1$ for all $0 < \tau \leq \|T^*T\|$;

(3) There is a constant $\gamma_* \geq 1$ such that $\sup_{0 \leq t \leq \|T^*T\|} \tau\, |g_\tau(t)| \leq \gamma_*$ for all $0 < \tau < \infty$,

where $r_\tau(t) := 1 - t g_\tau(t),\ 0 \leq t \leq \|T^*T\|$, denotes the residual function.

In regularization theory, the parametric family g_τ is usually parametrized with some parameter α as g_α. Since this symbol is used to note the bound for the error of the first kind, we instead use the symbol τ, contrasting to the usual convention.

Having chosen a specific regularization scheme g_τ we assign as reconstruction the linear mapping $R_\tau := g_\tau(T^*T)T^* \colon K \to H$. Notice that now, the element f_R is obtained as $f_R = f_\tau = g_\tau(T^*T)T^*Tf$. For the sake of convenience, we will write

$$\Phi_{\alpha,\tau} := \Phi_{\alpha,R_\tau}. \tag{4.3}$$

Example 2 (Truncated svd, spectral cut-off)**.** With this new notation we can use the function $g_\tau(t) := 1/t,\ t \geq \tau$, and zero elsewhere. This means that we approximate the inverse mapping of T as in Example 1. The condition $s_j^2 \geq \tau$ translates into an upper bound $1 \leq j \leq D = D(\tau)$. The element f_τ is then given as $f_\tau = \sum_{j=1}^{D} \langle f, u_j \rangle_H u_j$.

Example 3 (Tikhonov regularization). Another common linear regularization scheme is given by $g_\tau(t) = 1/(t+\tau)$, $t, \tau > 0$. In this case we have $R_\tau Y = (\tau I + T^*T)^{-1} T^*Y$, i.e., this is the minimizer of the penalized least squares functional $J_\tau(f) := \|Y - Tf\|_K^2 + \tau\|f\|_H^2$, $f \in H$.

Having chosen any linear regularization, we would like to bound the quantities $S_\tau^2 = S_R^2, v_\tau^2 = v_R^2$ from (2.4), (2.5) (with a slight abuse of notation). Under the above assumption, the reconstructions R_τ are also Hilbert–Schmidt operators, since these are compositions involving T^*.

In the following, we shall use the *effective dimension* which allows us to construct a bound on the variance S_τ^2.

Definition 2 (Effective dimension, see [6, 26]). The function $\lambda \mapsto \mathcal{N}(\lambda)$ defined as

$$\mathcal{N}(\lambda) := \operatorname{tr}\left[(T^*T + \lambda I)^{-1} T^*T\right] \tag{4.4}$$

is called effective dimension of the operator T^*T under white noise.

By Assumption A3 the operator T^*T has a finite trace, and the operator $(T^*T + \lambda I)^{-1}$ is bounded, thus the function $\mathcal{N}$ is finite. The following bound is a consequence of [4, Lem. 3.1]:

$$\operatorname{tr}\left[g_\tau^2(T^*T)T^*T\right] \le 2\gamma_*^2 \frac{\mathcal{N}(\tau)}{\tau}, \tag{4.5}$$

for some constant $\gamma_* > 0$. This, and using the definition of regularization schemes, results in the following bounds.

Lemma 4.1. *Let $R_\tau := g_\tau(T^*T)T^* \colon K \to H$. Assume that Assumption* A2 *holds, then we have*

(i) $S_\tau^2 \le 2\gamma_*^2\sigma^2 \dfrac{\mathcal{N}(\tau)}{\tau}$, $\tau > 0$, *and*

(ii) $v_\tau^2 \le \gamma_*^2\sigma^2 \dfrac{1}{\tau}$, $\tau > 0$.

Proof. The proof is a direct consequence of the definition of S_τ^2, v_τ^2 and of (4.5). □

This lemma provides only upper bounds for S_τ and v_τ. For many linear regularization schemes we can actually show that $v_\tau/S_\tau \to 0$ as $\tau \to 0$, and we mention the following result, whose proof is postponed to Appendix A.

Lemma 4.2. *Suppose that the regularization g_τ has the following properties:*

(1) *there are constants $\hat{c}, \hat{\gamma} > 0$ such that $|g_\tau(\hat{c}\tau)| \ge \hat{\gamma}/\tau$ for $\alpha > 0$, and*

(2) *for each $0 < t \le \|T^*T\|$ the function $\tau \to |g_\tau(t)|$ is decreasing.*

If the singular numbers of the operator T decay moderately, so that $\#\{j,\ \hat{c}\tau \le s_j^2 \le \hat{c}/\underline{c}\tau\} \to \infty$ as $\tau \to 0$, then $\operatorname{tr}\left[\tau g_\tau^2(T^*T)T^*T\right] \to \infty$ as $\tau \to 0$. Consequently, in this case we have $v_\tau/S_\tau \to 0$ as $\tau \to 0$ and

$$\frac{r^2(\Phi_{\alpha,\tau}, \beta)}{C_{\alpha,\beta}^* S_\tau v_\tau} \to 1 \quad \text{as } \tau \to 0,$$

where $C_{\alpha,\beta}^*$ and $r^2(\Phi_{\alpha,\tau}, \beta)$ are as in (3.7) *and* (3.8)*, respectively.*

Remark 4.1. The assumptions imposed above on g_τ are known to hold for many regularization schemes, in particular, for spectral cut-off and (iterated) Tikhonov regularization. The assumption on the singular numbers holds for (at most) polynomial decay.

We turn to bounding the bias $\|f - f_\tau\|_H$. This can be done under the assumption that the chosen regularization has a certain qualification, see, e.g., [11]. The concept of qualification is made precise in the following definition.

Definition 3 (Qualification)**.** Suppose that φ is an index function as in Assumption A4. The regularization g_τ is said to have qualification φ if there is a constant $\gamma < \infty$ such that

$$\sup_{0\le t\le \|T^*T\|} |r_\tau(t)|\,\varphi(t) \le \gamma\varphi(\tau), \quad \tau > 0.$$

Remark 4.2. It is well known that Tikhonov regularization has qualification $\varphi(t) = t$ with constant $\gamma = 1$, and this is the maximal power. On the other hand, every index function is a qualification with constant $\gamma = 1$ of truncated svd.

The concept of a qualification can be used to bound the bias at the element f_τ.

Proposition 4.1. *Let g_τ be any regularization having qualification φ with constant γ. If $f \in \mathcal{E}_\varphi$ then*

$$\|f - f_\tau\|_H \le \gamma\varphi(\tau).$$

Proof. Let ω with $\|\omega\|_H \le 1$ be such that $f = \varphi(T^*T)\omega$. Then

$$\|f_\tau - f\|_H = \|g_\tau(T^*T)T^*Tf - f\|_H = \|r_\tau(T^*T)f\|_H = \|r_\tau(T^*T)\varphi(T^*T)\omega\|_H \le \gamma\varphi(\tau).$$

□

Now we have established bounds for all quantities occurring in (4.2), and this yields the main result for linear regularization.

Theorem 4.1. *Assume that Assumption* A3 *holds, and suppose that g_τ is a regularization which has qualification φ, and that $f \in \mathcal{E}_\varphi$. Let τ_* be chosen from the equation*

$$\varphi^2(\tau) = \sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau}. \tag{4.6}$$

Then, for all $f \in \mathcal{E}_\varphi$,

$$\inf_{\tau>0} \left(r^2(\Phi_{\alpha,\tau},\beta) + \|f - f_\tau\|_H^2\right) \le \Big(C^*_{\alpha,\beta}\sqrt{2}\gamma_*^2 + \frac{(4x_\alpha + 8x_\beta)\gamma_*^2}{\sqrt{\mathcal{N}(\tau_*)}} + \gamma^2\Big)\varphi^2(\tau_*),$$

where the constant $C^*_{\alpha,\beta}$ has been introduced in (3.7)*. In particular, we get that*

$$\rho^2(\Phi_{\alpha,\tau_*},\beta,\mathcal{E}_\varphi) \le 2\Big(C^*_{\alpha,\beta}\sqrt{2}\gamma_*^2 + \frac{(4x_\alpha + 8x_\beta)\gamma_*^2}{\sqrt{\mathcal{N}(\tau_*)}} + \gamma^2\Big)\varphi^2(\tau_*).$$

Proof. By Propositions 4.1 and 3.1, we have

$$\begin{aligned} r^2(\Phi_{\alpha,\tau},\beta) + \|f - f_\tau\|_H^2 &= C^*_{\alpha,\beta}Sv + (4x_\alpha + 8x_\beta)v^2 + \|f - f_\tau\|_H^2 \\ &\le C^*_{\alpha,\beta}\sqrt{2}\gamma_*^2\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + (4x_\alpha + 8x_\beta)\gamma_*^2\sigma^2\frac{1}{\tau} + \gamma^2\varphi^2(\tau) \\ &\le \Big(C^*_{\alpha,\beta}\sqrt{2}\gamma_*^2 + \frac{(4x_\alpha + 8x_\beta)\gamma_*^2}{\sqrt{\mathcal{N}(\tau_*)}} + \gamma^2\Big)\varphi^2(\tau_*), \end{aligned}$$

since the parameter τ_* equates both terms $\varphi^2(\tau)$ and $\sigma^2\tau^{-1}\sqrt{\mathcal{N}(\tau)}$. This gives the upper bound. □

Remark 4.3. Up to now, all the presented results are non-asymptotic in the sense that we do not require that $\sigma^2 \to 0$. In an asymptotic setting, we can remark that τ_* as defined in (4.6) satisfies $\tau_* \to$ as $\sigma \to 0$. Since the effective dimension tends to infinity as $\tau \to 0$, we get that

$$\rho^2(\Phi_{\alpha,\tau_*}, \beta, \mathcal{E}_\varphi) \leq 2\big(C^*_{\alpha,\beta}\sqrt{2}\gamma_*^2(1 + o(1)) + \gamma^2\big)\varphi^2(\tau_*),$$

as $\sigma \to 0$.

We shall highlight the above results with two examples. We shall dwell into these in order to show that the above results are consistent with other results for inverse testing (see for instance [18]).

Example 4 (Moderately ill-posed problem)**.** Let us assume that the singular numbers of the operator T decay as $s_k \asymp k^{-t},\ k \in \mathbb{N}^*$, with $t > 1/2$ (in order to ensure that Assumption A3 is satisfied). In this case the effective dimension asymptotically behaves like $\mathcal{N}(\tau) \asymp \tau^{-1/(2t)}$, as $\tau \to 0$, see for instance [4, Ex. 3]. The Sobolev ball

$$\mathcal{E}^{\mathcal{X}}_{a,2} := \Big\{ f,\ \sum_{j=1}^{\infty} a_j^2 \langle f, \phi_j^2 \rangle_H \leq R^2 \Big\}, \quad \text{with } a_j = j^s, \quad \forall j > 1, \tag{4.7}$$

as considered in [18] coincides (up to constants) with $\mathcal{E}_\varphi$ for the function $\varphi(u) = u^{s/(2t)}$, $u > 0$. In this case the value τ_* from (4.6) is computed as $\tau_* \asymp \sigma^{8t/(4s+4t+1)}$, which results in an asymptotic separation rate of

$$\rho(\Phi_{\alpha,\tau_*}, \beta, \mathcal{E}_\varphi) \asymp \varphi(\tau_*) \asymp \sigma^{2s/(2s+2t+1/2)}, \qquad \sigma \to 0,$$

which corresponds to the 'mildly ill-posed case' in [18] or [15], and it is known to be minimax.

Example 5 (Severely ill-posed problem)**.** Here we assume a decay of the singular numbers of the form $s_k \asymp \exp(-\gamma k)$, $k \in \mathbb{N}$. The effective dimension behaves like $\mathcal{N}(\tau) \asymp \frac{1}{\gamma}\log(1/\tau)$. The Sobolev ball from (4.7) is now given as $\mathcal{E}_\varphi$ for a function $\varphi(u) = \left(\frac{1}{2\gamma}\log(1/u)\right)^{-s}$. Then the value τ_* calculates as $\tau_* \asymp \sigma^2\left(\log(1/\sigma^2)\right)^{2s+1/2}$, which results in a separation rate

$$\rho(\Phi_{\alpha,\tau_*}, \beta, \mathcal{E}_\varphi) \asymp \varphi(\tau_*) \asymp \log^{-s}(1/\sigma^2), \qquad \sigma \to 0,$$

again recovering the corresponding result from [18].

4.2. Projection Schemes

Details on the solution of ill-posed equations by using projection schemes can be found in [21, 23, 25], and our outline follows the recent [21]. In particular we use the intrinsic requirements such as quasi-optimality and robustness of projection schemes in order to obtain a control similar to the previous section.

We fix a finite-dimensional subspace $H_m \subset H$, called the *design space* and/or a finite-dimensional subspace $K_n \subset K$, called the *data space*. We shall denote throughout the corresponding orthogonal projections onto H_m by P_m, and/or the orthogonal projection onto K_n by Q_n. The subscripts m and n denote the *dimensions* of the spaces. Given such couple (H_m, K_n) of spaces we turn from the equation (1.1) to its discretization

$$Q_n Y = Q_n T P_m x + \sigma Q_n \xi. \tag{4.8}$$

Equation (4.8) corresponds to a discretization as follows. If $\psi_1, \ldots, \psi_n$ denotes a basis of K_n then we retrieve as data the coefficients $\langle Y, \psi_j \rangle$, $j = 1, \ldots, n$, of Y in this basis. Also, if $\varphi_1, \ldots, \varphi_m$ is a basis for H_m then we aim at representing the unknown solution element in a corresponding series expansion. This results in an $n \times m$ system of linear equations with matrix M having entries $M_{j,k} = \langle T\varphi_j, \psi_k \rangle$.

Without further assumptions, the finite-dimensional equation (4.8) may have no or many solutions, and hence we shall turn to the *least-squares solution* as given by the *Moore–Penrose* inverse, i.e., we assign

$$f_{m,n} := (Q_n T P_m)^\dagger Q_n Y. \tag{4.9}$$

Definition 4 (Projection scheme, see [21]). If we are given

(1) an increasing sequence $H_1 \subset H_2 \cdots \subset H$, and

(2) an increasing sequence $K_1 \subset K_2 \cdots \subset K$, together with

(3) a mapping $m \to n(m),\ m = 1, 2, \dots,$

then the corresponding sequence of mappings

$$Y \to f_{m,n(m)} := (Q_n T P_m)^{\dagger} Y \tag{4.10}$$

is called a *projection scheme*.

Example 6 (Truncated svd, spectral cut-off). The truncated svd, as introduced in Example 1 is also an example for a projection scheme, if we use the increasing sequences $H_m := \operatorname{span}\{u_1, \dots, u_m\} \subset H$, and $K_m := \operatorname{span}\{v_1, \dots, v_m\} \subset K$, respectively. In this case we see that

$$(Q_n T P_m)^{\dagger} Y = \sum_{j=1}^{m} \frac{1}{s_j} \langle Y, v_j \rangle u_j.$$

Henceforth we shall always assume that the mapping $(Q_n T P_m)^{\dagger} \colon K_n \to H_m$ is invertible, i.e., the related linear system of equations has a unique solution. This gives an (implicit) relation $n = n(m)$, typically $n = m$ will do. However, our subsequent analysis will be done using the dimension m of the space H_m. In accordance with this we will denote f_R by f_m, highlighting the dependence on the dimension. Thus the linear reconstruction R is given as $R := (Q_n T P_m)^{\dagger}$, and we define

$$\Phi_{\alpha,m} \colon = \Phi_{\alpha,R}. \tag{4.11}$$

We need to control $\operatorname{tr}[R^* R]$ as well as $\|R\|$. The latter is related to the robustness (stability) of the projection scheme.

Definition 5 (Robustness). A projection scheme $\left((Q_n T P_m)^{\dagger},\ m \in \mathbb{N}\right)$ is said to be *robust* if there is a constant $D_R < \infty$ for which

$$\|(Q_n T P_m)^{\dagger}\| \le \frac{D_R}{\mathrm{j}(T, H_m)}, \qquad m = 1, 2, \dots \tag{4.12}$$

Here, the quantity $\mathrm{j}(T, H_m)$ denotes the *modulus of injectivity* of T with respect to the subspace H_m, given as

$$\mathrm{j}(T, H_m) := \inf_{0 \neq x \in H_m} \frac{\|Tx\|_K}{\|x\|_H}. \tag{4.13}$$

The modulus of injectivity is always smaller than the m-th singular number $s_m = s_m(T)$ of the mapping T, and hence we say that the subspaces H_m satisfy a *Bernstein-type* inequality if there is a constant $0 < C_B \le 1$ such that

$$C_B s_m(T) \le \mathrm{j}(T, H_m).$$

We summarize our previous considerations as follows.

Lemma 4.3. *Suppose that the projection scheme $\left((Q_n T P_m)^{\dagger},\ m \in \mathbb{N}\right)$ is robust and that the spaces H_m obey a Bernstein-type inequality. Then*

$$\|(Q_n T P_m{}^{\dagger})\| \le \frac{D_R}{C_B} \frac{1}{s_m}.$$

In particular we have

$$v_R^2 := v_m^2 \le \sigma^2 \frac{D_R^2}{C_B^2} \frac{1}{s_m^2}.$$

We turn to bounding S_R^2. Before doing so we mention that for spectral cut-off from Example 6, this bound can easily be established.

Lemma 4.4. *For spectral cut-off we have*

$$S_R^2 = \sigma^2 \operatorname{tr}\left[\left((Q_nTP_m)^\dagger\right)^*(Q_nTP_m)^\dagger\right] = \sigma^2 \sum_{j=1}^m \frac{1}{s_j^2}.$$

In order to obtain a similar bound in more general situations we need to impose restrictions on the decay of the singular numbers s_j, $j = 1, 2, \dots$

The use of projection schemes for severely ill-posed problems requires a particular care, and the following restriction, which will be imposed on the decay of the singular numbers of the operator T takes this into account. We shall assume that the decreasing sequence s_j, $j = 1, 2, \dots$, is *regularly varying* for some index $-r \le 0$, and we refer to [5] for a treatment. In this case the corresponding sequence s_j^{-2}, $j = 1, 2, \dots$, is regularly varying with index $2r$, and we have

$$\frac{1}{m} s_m^2 \sum_{j=1}^m \frac{1}{s_j^2} \longrightarrow \frac{1}{2r+1} \quad \text{as } m \to \infty.$$

In particular, there is a constant C_r such that

$$\frac{m}{s_m^2} \le C_r^2 \sum_{j=1}^m \frac{1}{s_j^2}, \tag{4.14}$$

and the latter bound is actually all that is needed.

Example 7 (Moderately ill-posed problems)**.** The special case where the singular numbers decay as $s_j \asymp j^{-r}$, $j = 1, 2, \dots$, is covered by the concept of regularly varying sequences, and such sequences have the index $-r$. In particular, for moderately ill-posed problems a bound (4.14) is valid.

Lemma 4.5. *Suppose that the sequence s_j, $j = 1, 2, \dots$, is such that* (4.14) *holds for a constant C_r. If the projection scheme is robust with constant D_R, and if the spaces H_n obey a Bernstein-type inequality with constant C_B then*

$$S_R^2 := S_m^2 = \sigma^2 \operatorname{tr}\left[\left((Q_nTP_m)^\dagger\right)^*(Q_nTP_m)^\dagger\right] \le 2C_r^2 \frac{D_R^2}{C_B^2} \sigma^2 \sum_{j=1}^m \frac{1}{s_j^2}.$$

If, in addition, Assumption A3 *is satisfied, then we have*

$$S_R^2 := S_m^2 \le C_r^2 \frac{D_R^2}{C_B^2} \sigma^2 \frac{\mathcal{N}(s_m^2)}{s_m^2}.$$

Proof. We notice that the mapping $\left((Q_nTP_m)^\dagger\right)^*$ is zero on $H_m^\perp$, the orthogonal complement of H_m. So, we take an orthonormal system $u_1, u_2, \dots, u_m, \dots$, where the first m components span H_m. With respect to this system we see that

$$\begin{aligned}
\operatorname{tr}\left[\left((Q_nTP_m)^\dagger\right)^*(Q_nTP_m)^\dagger\right] &= \operatorname{tr}\left[(Q_nTP_m)^\dagger\left((Q_nTP_m)^\dagger\right)^*\right] \\
&= \sum_{j=1}^\infty \left\|\left((Q_nTP_m)^\dagger\right)^* u_j\right\|_K^2 = \sum_{j=1}^m \left\|\left((Q_nTP_m)^\dagger\right)^* u_j\right\|_K^2 \\
&\le m\left\|\left((Q_nTP_m)^\dagger\right)^*\right\|^2 = m\left\|(Q_nTP_m)^\dagger\right\|^2.
\end{aligned}$$

Using Lemma 4.3 we see that

$$\operatorname{tr}\left[\left((Q_nTP_m)^\dagger\right)^*(Q_nTP_m)^\dagger\right] \le m \frac{D_R^2}{C_B^2} \frac{1}{s_m^2}.$$

Now we use (4.14) to complete the proof of the first assertion. Under Assumption A3 we continue and use the inequality $u/v \le 2v/(u+v)$, $0 < u \le v$, to see that

$$\sum_{j=1}^{m} \frac{1}{s_j^2} \le \frac{2}{s_m^2} \sum_{j=1}^{m} \frac{s_j^2}{s_j^2 + s_m^2} \le 2\frac{\mathcal{N}(s_m^2)}{s_m^2},$$

and the proof is complete. □

Remark 4.4. Notice that Lemma 4.5 provides us with (an order optimal) bound for the variance, even if the operator T is not a Hilbert–Schmidt one. But, if it is, then the obtained bound corresponds to the one from Lemma 4.1 (with $\tau \leftarrow s_m^2$).

Next, we need to bound $\|f - f_R\|_H$, as was done in § 4.1 by assuming qualification, and we need a further property of the projection scheme called *quasi-optimality*. We start with the following well-known result, originally from spline interpolation [8], and used for projection schemes in [23], which states that

$$\|f - (Q_n T P_m)^{\dagger} T f\|_H \le \|(Q_n T P_m)^{\dagger} T\| \, \|f - P_m f\|_H. \tag{4.15}$$

Therefore, we can bound the bias on the left whenever the norms $\| (Q_n T P_m)^{\dagger} T\|$ are uniformly bounded.

Definition 6 (Quasi-optimality). A projection scheme $Y \to (Q_n T P_m)^{\dagger} Y$ is *quasi-optimal* if there is a constant D_Q such that $\|(Q_n T P_m)^{\dagger} T\| \le D_Q$.

We emphasize that under quasi-optimality the bound for the bias entirely depends on the approximation power of the projections P_m with respect to the element f. This approximation power is expressed in terms of degree of approximation, whose definition is made explicit below.

Definition 7 (Degree of approximation). *Suppose that* $\{H_m\}$, $\dim(H_m) \le m$, *is a nested set of design spaces. The spaces* H_m *are said to have the* degree of approximation φ *it there is a constant* $C_D < \infty$ *with*

$$\|(I - P_m)\varphi(T^*T)\| \le C_D \varphi(s_{m+1}), \quad m = 1, 2, \ldots \tag{4.16}$$

For spectral cut-off this bound (with constant $C = 1$) is best possible. Also, using interpolation type inequalities one can verify this property for many known approximation spaces H_m, $m = 1, 2, \ldots$, we refer to [21] for more details on degree of approximation and Bernstein-type bounds. We now can state the analogue of Proposition 4.1 for projection schemes.

Proposition 4.2. *Suppose that the projection scheme is quasi-optimal with constant* D_Q, *and that it has the degree of approximation* φ *with constant* C_D. *If* $f \in \mathcal{E}_{\varphi}$ *then we have*

$$\|f - f_m\|_H \le D_Q C_D \varphi(s_{m+1}^2).$$

We return to the optimization problem raised in (4.2). Here, the family of reconstructions R runs over all projection schemes, and we can control the bound by a proper choice of the discretization level m.

For the sake of convenience, we will assume in the following that Assumption A3 is satisfied, i.e., that T is a Hilbert–Schmidt operator. If it is not the case, Theorem 4.2 below remains valid when replacing $\sqrt{\mathcal{N}(s_m^2)}/s_m^2$ by $\sqrt{\sum_{j=1}^{m} s_j^{-2}}/s_m$.

Theorem 4.2. *Suppose that the approximate solutions are obtained by a projection scheme which is quasi-optimal and robust and that Assumption* A3 *holds. Furthermore assume that the design spaces* H_m *have degree of approximation* φ *and obey a Bernstein-type inequality. Let* m_* *be chosen from*

$$m_* = \max\left\{m, \quad \varphi^2(s_m^2) \ge \sigma^2 \frac{\sqrt{\mathcal{N}(s_m^2)}}{s_m^2}\right\}. \tag{4.17}$$

If $f \in \mathcal{E}_\varphi$ then we have

$$\inf_m \{r^2(\Phi_{\alpha,m},\beta) + \|f - f_m\|_H^2\} \le \left(C^*_{\alpha,\beta}\frac{D_R^2}{C_B^2}C_r + (4x_\alpha + 8x_\beta)\frac{D_R^2}{C_B^2}\frac{1}{\sqrt{\mathcal{N}(s^2_{m_*})}} + D_Q^2C_D^2\right)\varphi^2(s^2_{m_*}),$$

where the constant $C^*_{\alpha,\beta}$ has been introduced in (3.7).

Proof. By using Lemma 4.5 and Proposition 4.2 we see that for any choice of discretization level m we have

$$\begin{aligned} r^2(\Phi_{\alpha,m},\beta) + \|f - f_m\|_H^2 &\le C^*_{\alpha,\beta}\frac{D_R^2}{C_B^2}C_r\sigma^2\frac{\sqrt{\mathcal{N}(s_m^2)}}{s_m^2} + (4x_\alpha + 8x_\beta)\sigma^2\frac{D_R^2}{C_B^2}\frac{1}{s_m^2} + D_Q^2C_D^2\varphi^2(s^2_{m+1}) \\ &\le \left(C^*_{\alpha,\beta}\frac{D_R^2}{C_B^2}C_r + (4x_\alpha + 8x_\beta)\frac{D_R^2}{C_B^2}\frac{1}{\sqrt{\mathcal{N}(s^2_{m_*})}} + D_Q^2C_D^2\right) \\ &\qquad \times \max\left\{\sigma^2\frac{\sqrt{\mathcal{N}(s_m^2)}}{s_m^2}, \varphi^2(s^2_{m+1})\right\}. \end{aligned}$$

At the discretization level $m_* + 1$ we see by monotonicity that

$$\varphi^2(s^2_{m_*+1}) \le \varphi^2(s^2_{m_*}).$$

Also, by the choice of m_* we see that

$$\sigma^2\frac{\sqrt{\mathcal{N}(s^2_{m_*})}}{s^2_{m_*}} \le \varphi^2(s^2_{m_*}),$$

hence both terms in the max are dominated by $\varphi^2(s^2_{m_*})$, which allows us to complete the proof. □

Once again, the previous result is non-asymptotic. In the asymptotic regime, we get the following improvement.

Corollary 4.1. *Under the assumptions of Theorem* 4.2 *we get that*

$$\inf_m \{r^2(\Phi_{\alpha,m},\beta) + \|f - f_m\|_H^2\} \le \left(C^*_{\alpha,\beta}\frac{D_R^2}{C_B^2}C_r(1 + o(1)) + D_Q^2C_D^2\right)\varphi^2(s^2_{m_*}),$$

as $\sigma \to 0$.

This is an easy consequence of the fact that along with $\sigma \to 0$ we have $s^2_{m_*} \to 0$, and hence the effective dimension at $s^2_{m_*}$ tends to infinity.

4.3. Discussion

We discuss the relation between nonparametric testing and function estimation. If Assumption A3 is satisfied then for both linear regularization and projection schemes, the optimal parameter τ_* ($\sim s^2_{m_*}$) is obtained by solving the "equation"

$$\sigma^2 = \tau\varphi^2(\tau)/\sqrt{\mathcal{N}(\tau)}.$$

This equation is different from the one which is to be solved for function estimation, given by

$$\sigma^2 = \tau\varphi^2(\tau)/\mathcal{N}(\tau).$$

The effective dimension $\mathcal{N}$, which is designed for estimation, enters in the inverse testing problem in square root, so that, loosely speaking, *testing is easier*.

Another remark may be of interest. For the estimation problem, the bias–variance decomposition yields the minimization problem

$$\|f - f_R\|_H^2 + S_R^2 \longrightarrow \text{MIN}$$

by a proper choice of reconstruction R. For testing, and we refer to (4.2), the corresponding minimization problem is

$$\|f - f_R\|_H^2 + S_R v_R \longrightarrow \text{MIN}.$$

Since, as already mentioned, $S_R v_R \leq S_R^2$, this calibration always yields a smaller value, which again explains the different rates for separation radius and estimation error.

5. RELATING THE DIRECT AND INVERSE TESTING PROBLEMS

For injective linear operators T, the assertions "$f = 0$" and "$Tf = 0$" are equivalent. Hence, testing $H_0\colon f = 0$ or testing $H_0\colon Tf = 0$ are related to the same problem: we want to detect whether there is signal in the data. Nevertheless, these testing problems are different in the sense that the alternatives are not expressed in the same way. Indeed, the inverse testing problem (considered in the previous sections) corresponds to

$$H_0^I\colon f = 0, \text{ against } H_1^I\colon f \in \mathcal{E}_\varphi,\ \|f\|_H^2 \geq (\rho^I)^2, \tag{5.1}$$

while the direct testing problem corresponds to testing

$$H_0^D\colon Tf = 0, \text{ against } H_1^D\colon f \in \mathcal{E}_\varphi,\ \|Tf\|_K^2 \geq (\rho^D)^2. \tag{5.2}$$

In this section, we investigate the similarities between these two view points. In particular, we remark that the two testing problems are not equivalent in the sense that the alternatives do not deal with the same object.

5.1. Relating the Separation Rates

The authors in [17] discussed whether both problems (5.1) and (5.2) are related. The main result, Theorem 1, ibid. asserts that for a variety of cases each minimax test Φ_α for the direct problem ($H_0\colon Tf = 0$) is also minimax for the related inverse problem ($H_0\colon f = 0$). This fundamental result is based on Lemma 1, ibid. Here we show that this lemma has its origin in *interpolation* in variable Hilbert scales, and we refer to [22]. Actually we do not need the machinery as developed there, but we may use the following special case, which may directly be proved using Jensen's inequality.

Lemma 5.1 (Interpolation inequality). *Let φ be from* (2.1) *in Assumption* A4*, and let $\Theta(u) := \sqrt{u}\varphi(u)$, $u > 0$. If the function $u \mapsto \varphi^2\big((\Theta^2)^{-1}(u)\big)$ is concave then*

$$\|f\|_H \leq \varphi\left(\Theta^{-1}(\|Tf\|_K)\right), \qquad f \in \mathcal{E}_\varphi. \tag{5.3}$$

The main result relating the direct and inverse testing problems is the following.

Theorem 5.1. *Let φ be an index function with related function Θ, such that the function $u \mapsto \varphi^2\big((\Theta^2)^{-1}(u)\big)$ is concave. Let Φ_α be a level-α test for the direct problem $H_0^D\colon Tf = 0$ with uniform separation rate $\rho^D(\Phi_\alpha, \beta, \mathcal{E}_\Theta)$. Then Φ_α constitutes a level-α test for the inverse problem $H_0^I\colon f = 0$ with uniform separation rate*

$$\rho^I(\Phi_\alpha, \beta, \mathcal{E}_\varphi) \leq \varphi\big(\Theta^{-1}(\rho^D(\Phi_\alpha, \beta, \mathcal{E}_\Theta))\big).$$

Consequently we have for the minimax separation rates that

$$\rho^I(\alpha, \beta, \mathcal{E}_\varphi) \leq \varphi\big(\Theta^{-1}(\rho^D(\alpha, \beta, \mathcal{E}_\Theta))\big). \tag{5.4}$$

Proof. Clearly, the test Φ_α is a level-α test for both problems, and we need to control the error of the second kind. But if $\|f\|_H \geq \varphi\big(\Theta^{-1}(\rho^D(\Phi_\alpha, \beta, \mathcal{E}_\Theta))\big)$ then Lemma 5.1 yields that $\|Tf\|_K \geq \rho^D(\Phi_\alpha, \beta, \mathcal{E}_\Theta)$, and the assertion is a consequence of the properties of the test for the direct problem.

If Φ_α was minimax for the direct problem then the corresponding minimax rate for the inverse problem must be dominated by $\varphi\big(\Theta^{-1}(\rho^D(\alpha, \beta, \mathcal{E}_\Theta))\big)$, which gives (5.4). ☐

Remark 5.1. In many cases the bound (5.4) actually is an asymptotic equivalence

$$\varphi^{-1}\left(\rho^I(\alpha, \beta, \mathcal{E}_\varphi)\right) \asymp \Theta^{-1}\left(\rho^D(\alpha, \beta, \mathcal{E}_\Theta)\right), \quad \sigma \to 0. \tag{5.5}$$

It may be enlightening to see this on the basis of Example 4. Recall that the function φ was given as $\varphi(u) = u^{s/(2t)}$. The corresponding rate is known to be minimax, and we obtain that

$$\varphi^{-1}\left(\rho^I(\alpha, \beta, \mathcal{E}_\varphi)\right) \asymp \sigma^{\frac{4t}{2s+2t+1/2}}.$$

We turn to the direct problem, for which the corresponding smoothness class is $\mathcal{E}_\Theta$ for the function $\Theta(u) = u^{2/(2t)+1/2} = u^{(s+t)/(st)}$. This corresponds to $\mu = s + t$ in [17, Tbl. 2], yielding the separation rate $\rho(\alpha, \beta, \mathcal{E}_\Theta) \asymp \sigma^{2(s+t)/(2s+2t+1/2)}$, which in turn gives

$$\Theta^{-1}\left(\rho^D(\alpha, \beta, \mathcal{E}_\Theta)\right) \asymp \sigma^{\frac{4t}{2s+2t+1/2}},$$

and hence (5.5) for moderately ill-posed problems.

Similarly, this holds for severely ill-posed problems, and we omit details.

We emphasize that, by virtue of Theorem 5.1, any lower bound for the minimax separation rate in the inverse testing problem yields a lower bound for the corresponding direct problem.

Remark 5.2. Thanks to Theorem 5.1, it is possible to prove that in all the cases considered in this paper, a test minimax for (5.2) will be also minimax for (5.1). Nevertheless, the reverse is not true. We will not dwell into details, instead we refer to [17] for a detailed discussion on this subject.

5.2. Designing Tests for the Direct Problem

The coincidence in (5.5) is not by chance and we indicate a further result in this direction. Recall from § 4.1 that the value of $\tau_* = \tau_*^{\mathrm{IP}}$ was obtained from (4.6), and hence that we actually have $\rho(\alpha, \beta, \mathcal{E}_\varphi) \asymp \varphi(\tau_*^{\mathrm{IP}})$, such that the left-hand side in (5.5) equals τ_*^{IP}. We shall see next that the corresponding value $\tau_* = \tau_*^{\mathrm{DP}}$ is obtained from the same equation (4.6) when basing the direct test on the family $\widehat{TR_\tau} = TR_\tau$ with family $R_\tau = g_\tau(T^*T)T^*$ as in § 4.1. Then $TR_\tau = g_\tau(TT^*)TT^*$, and we bound its variance and weak variance, next.

Lemma 5.2. *Let $\tilde{R}_\tau = TR_\tau = g_\tau(TT^*)TT^*$ and denote by resp. $\tilde{S}_\tau^2$ and $\tilde{v}_\tau^2$ the corresponding strong and weak variance. If Assumption* A3 *holds then*

(1) $\tilde{S}_\tau^2 \leq (\gamma_0 + \gamma_*)\gamma_0\sigma^2\mathcal{N}(\tau)$, $\tau > 0$, *and*

(2) $\tilde{v}_\tau^2 \leq \sigma^2\gamma_0^2$.

We also need to bound the bias $\|Tf - Tf_\tau\|_K$ with $f_\tau = g_\tau(T^*T)T^*Tf$.

Lemma 5.3. *Assume that $f \in \mathcal{E}_\varphi$. If the regularization g_τ has qualification Θ with constant γ then*

$$\|Tf - Tf_\tau\|_K \leq \gamma\Theta(\tau).$$

Proof. Since $f_\tau = R_\tau Tf$, we get that

$$\|Tf - Tf_\tau\|_K = \|Tf - g_\tau(TT^*)TT^*Tf\|_K = \|r_\tau(TT^*)Tf\|_K,$$

which is bounded by $\gamma\Theta(\tau)$ as soon as $f \in \mathcal{E}_\varphi$ and g_τ has qualification Θ. ☐

We recall from Section 4 the quantity $r^2(\Phi_\alpha, \beta) := C_{\alpha,\beta} S_\tau v_\tau$, where we now consider $\tilde{S}_\tau$ and $\tilde{v}_\tau$ from Lemma 5.2 for bounding $\|Tf\|_K^2 \geq C_{\alpha,\beta}\tilde{S}_\tau \tilde{v}_\tau$ from below.

Corollary 5.1. *Suppose that g_τ is a regularization which has qualification Θ, $f \in \mathcal{E}_\varphi$ and that Assumption* A3 *holds. Let τ_*^{DP} be chosen from the equation*

$$\sigma^2 = \frac{\Theta^2(\tau)}{\sqrt{\mathcal{N}(\tau)}}. \tag{5.6}$$

Then

$$\inf_{\tau > 0} \left(r^2(\Phi_\alpha, \beta) + \|Tf - Tf_\tau\|_K^2 \right) \leq \left(C_{\alpha,\beta}\sqrt{(\gamma_0 + \gamma_*)\gamma_0}\gamma_0 + \gamma^2 \right) \Theta^2(\tau_*^{\mathrm{DP}}).$$

We stress that the equation (5.6) for determining τ_*^{DP} is the same as (4.6), since $\Theta^2(\tau) = \tau\varphi^2(\tau)$, and this explains the identical asymptotics in (5.5) as being equal to $\tau_*^{\mathrm{DP}} = \tau_*^{\mathrm{IP}}$.

This result sheds light on another interesting problem: If we want to use the regularization TR_τ, and if we want to have this optimal performance properties then the underlying regularization g_τ must have *higher qualification* Θ for the direct problem as compared for its use in inverse testing requiring qualification φ, only. This cannot be seen when confining to spectral cut-off, but this problem is relevant when considering other regularization schemes for testing. It is thus interesting to design estimators for $g = Tf$ which do not rely on estimation of f. However, since the data Y do not belong to the space K either discretization or some other kind of preconditioning is necessary in order to estimate $g = Tf$ from the data Y. Such direct estimation is simple by using projection schemes, and we exhibit the calculus for one-sided discretization. As in § 4.2, we choose finite (m) dimensional subspaces $Y_m \subset K$, with corresponding projections Q_m and consider the data

$$Q_m Y = Q_m g + \sigma Q_m \xi, \quad m \in \mathbb{N}.$$

This approach is called *dual least squares* scheme in regularization, see [23]. Here it is easy to see that $S_m^2 = \operatorname{tr}[Q_m^* Q_m] = m$, while $v_m^2 = \|Q_m\|^2 = 1$. In order to continue we just need that the chosen projections have degree of approximation Θ, i.e, there is C_D for which $\|(I - Q_m)\Theta(TT^*)\|_K \leq C_D\Theta(s_{m+1}^2)$, $m = 1, 2\ldots$ With this requirement at hand we can continue as if the projections Q_m were the projections onto the first m singular elements in the svd of T. In particular we have the upper bound on the separation radius

$$\rho(\mathcal{E}_\Theta, \alpha, \beta) \leq \max\{C_{\alpha,\beta}, C_D^2\} \inf_m \left(\sigma^2\sqrt{m} + \Theta^2(s_{m+1}^2) \right),$$

similar to corresponding results obtained for spectral cut-off in [1, 17], and we omit further details.

6. ADAPTATION TO THE SMOOTHNESS OF THE ALTERNATIVE

It seems clear from Section 4 that the optimality of the considered tests strongly depends on the regularity (smoothness) of the alternative. In this section, we propose data-driven tests that automatically adapt to the unknown smoothness parameter. The adaptation issue in test theory has widely been investigated. For more details on the subject, we refer for instance to [2], [24] in the direct setting (i.e., $T = Id$) or [15] in the inverse case for an adaptive scheme based on the singular-value decomposition of the operator.

First, we propose a general adaptive scheme. Then, we apply this approach to linear regularization over ellipsoids. This methodology can also be extended to projection schemes. For the sake of brevity, this extension is not discussed here.

6.1. A General Scheme for Adaptation

Assume that we have at our disposal a finite collection $(R)_{R\in\mathcal{R}}$ of regularization operators satisfying Assumption A3. Then, we can associate with each operator R a level-α test $\Phi_{\alpha,R}$. Our aim in this section is to construct a test that mimics the behavior of the best possible test among the family $\mathcal{R}$. Let $|\mathcal{R}|$ denotes the cardinality of the family $\mathcal{R}$. We define our adaptive test $\Phi^\star_\alpha$ as

$$\Phi^\star_\alpha = \max_{R\in\mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|},R}. \tag{6.1}$$

The performance of $\Phi^\star_\alpha$ is summarized in the following proposition.

Proposition 6.1. *The test introduced in* (6.1) *is a level-α test. Moreover*

$$P_f(\Phi^\star_\alpha = 0) \le \beta,$$

as soon as

$$\|f\|^2_H \ge 2 \inf_{R\in\mathcal{R}} \big(r^2(\Phi_{\frac{\alpha}{|\mathcal{R}|},R},\beta) + \|f - f_R\|^2_H\big),$$

where the function r^2 has been introduced in (3.8).

Proof. We first remark that

$$P_{H_0}(\Phi^\star_\alpha = 1) = P_{H_0}\big(\max_{R\in\mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 1\big) = P_{H_0}\Big(\bigcup_{R\in\mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 1\Big)$$
$$\le \sum_{R\in\mathcal{R}} P_{H_0}\big(\Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 1\big) = \alpha,$$

since $P_{H_0}(\Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 1) = \alpha/|\mathcal{R}|$ for all $R \in \mathcal{R}$. Hence, $\Phi^\star_\alpha$ is a level-α test. Now, we can investigate the error of the second kind. Using simple algebra, we get that

$$P_f(\Phi^\star_\alpha = 0) = P_{H_0}\big(\max_{R\in\mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 0\big) = P_{H_0}\Big(\bigcap_{R\in\mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 0\Big)$$
$$\le \inf_{R\in\mathcal{R}} P_{H_0}\big(\Phi_{\frac{\alpha}{|\mathcal{R}|},R} = 0\big).$$

We can conclude using (4.1). □

Proposition 6.1 proves that the detection radius associated with the test defined in (6.1) is close to the smallest possible one among the family $\mathcal{R}$. Thus, we must design the set $\mathcal{R}$ according to two requirements. First, the cardinality $|\mathcal{R}|$ should be small, in order not to enlarge the detection radius too much. Indeed, the following holds true.

Lemma 6.1. *Let $C^*_{\alpha,\beta}$ be the term introduced in* (3.7). *If the family $\mathcal{R}$ of regularization schemes has cardinality $M := |\mathcal{R}| \ge 1$, then*

$$C^*_{\alpha/M,\beta} \le C^*_{\alpha,\beta} + 2\sqrt{2\log(M)}.$$

If $M \ge 4$ then $C^*_{\alpha/M,\beta} \le (C^*_{\alpha,\beta} + 2\sqrt{2})\sqrt{\log(M)}$.

Proof. We first observe that $x_{\alpha/M} = x_\alpha + \log(M)$. Therefore we conclude that

$$C^*_{\alpha,\beta} = 2\sqrt{x_\beta} + 2\sqrt{2x_{\alpha/M}}$$
$$= C^*_{\alpha,\beta} + 2\big(\sqrt{2x_\alpha + 2\log(M)} - \sqrt{2x_\alpha}\ \big) \le C^*_{\alpha,\beta} + 2\sqrt{2\log(M)}.$$

The second assertion is trivial, because $\log(M) > 1$ for $M \ge 4$. □

Therefore, the price to pay for using $\Phi_\alpha^\star$ is a term of order $\sqrt{\log(|\mathcal{R}|)}$, up to some condition on the behavior of the effective dimension (see Theorem 6.1 below). On the other hand, the set $\mathcal{R}$ should be rich enough to keep the detection radius on the size of the best possible bound, as was established in Theorems 4.1 and 4.2.

In the following, we propose practical situations where such an adaptive scheme can be used. In particular, we propose families of regularization operators with controlled size and prove that the adaptive test Φ_α^* attains the minimax rate of testing (up to a log log term) for a proper choice of $\mathcal{R}$.

Remark 6.1. In the test (6.1), each regularization operator $R \in \mathcal{R}$ is associated with a test $\Phi_{\frac{\alpha}{|\mathcal{R}|},R}$ having the same level $\alpha/|\mathcal{R}|$. It is nevertheless possible to use more refined approaches, leading to an improvement of the power of the test (in terms of the constants). We refer to [10, Eq. (2.2)], however in a slightly different setting.

6.2. Application to Linear Regularization

We will exhibit the use of the general methodology for tests based on linear regularization.

Let g_τ be a given regularization. We associate with each function g_τ the operator R_τ and we deal with the finite family $\mathcal{R} = (R_\tau)_{\tau\in\mathcal{M}}$, where $\mathcal{M} \subset (0,\infty)$ is a finite set. We apply Proposition 6.1 with the corresponding test Φ_α^*. To this end we will use an exponential grid. Given an initial value $\tau_{\max}$, and a tuning parameter $0 < q < 1$ we consider as set $\mathcal{M}$ the exponential grid

$$\Delta_q := \{\tau = q^j\tau_{\max},\ j = 0,\dots,M-1\}, \quad \text{for some } M > 1. \tag{6.2}$$

Then we use the adaptive test

$$\Phi_\alpha^\star = \max_{\tau\in\Delta_q} \Phi_{\alpha/M,\tau}, \tag{6.3}$$

where the tests $\Phi_{\cdot,\tau}$ have been introduced in (4.3). The result of Proposition 6.1 can be rephrased as follows. By virtue of Lemma 6.1, and using the bounds from Lemma 4.1 and Proposition 4.1, respectively, we find that the test Φ_α^* bounds the error of the second kind by β as soon as

$$\|f^2\|_H \geq C(\alpha,\beta) \inf_{\tau\in\Delta_q}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right),$$

for some explicit constant $C(\alpha,\beta)$. We shall now show, how we can specify the numbers $0 < \tau_{\min} < \tau_{\max}$ such that this is of the order of the separation radius (up to a log log-factor).

The cardinality M obeys $\tau_{\min} := q^{M-1}\tau_{\max}$, and hence $M := \log_{\frac{1}{q}}(\tau_{\max}/\tau_{\min})$. Obviously we have

$$\inf_{\tau_{\min}\leq\tau\leq\tau_{\max}}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right) \leq \inf_{\tau\in\Delta_q}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right).$$

The reverse is also true (up to some constant), as proved in the following lemma.

Lemma 6.2 (cf. [16, Proof of Thm. 3.1]). *We have*

$$\inf_{\tau_{\min}\leq\tau\leq\tau_{\max}}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right) \geq q^{3/2}\inf_{\tau\in\Delta_q}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right).$$

Proof. For any τ with $\tau_{\min} < \tau \le \tau_{\max}$ we find an index $1 \le j \le M$ for which $\tau_j < \tau \le \tau_j/q$. The crucial observation is that the function $\tau \to \frac{\sqrt{\mathcal{N}(\tau)}}{\tau}$ is decreasing, whereas the function $\tau \to \sqrt{\tau \mathcal{N}(\tau)} = \tau^{3/2} \frac{\sqrt{\mathcal{N}(\tau)}}{\tau}$ is increasing, which can be seen from spectral calculus. Therefore, by using the above monotonicity we see that

$$\begin{aligned}
&\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau) \\
&\quad \ge \sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_j/q)}}{\tau_j/q} + \log(M)\frac{\sigma^2}{\tau_j/q} + \varphi^2(\tau_j) \\
&\quad = \sqrt{\log(M)}\sigma^2 \left(\frac{\tau_j}{q}\right)^{-3/2} \left(\frac{\tau_j}{q}\right)^{3/2} \frac{\sqrt{\mathcal{N}(\tau_j/q)}}{\tau_j/q} + q\log(M)\frac{\sigma^2}{\tau_j} + \varphi^2(\tau_j) \\
&\quad \ge \sqrt{\log(M)}\sigma^2 \left(\frac{\tau_j}{q}\right)^{-3/2} {\tau_j}^{3/2} \frac{\sqrt{\mathcal{N}(\tau_j)}}{\tau_j} + q^{3/2}\log(M)\frac{\sigma^2}{\tau_j} + \varphi^2(\tau_j) \\
&\quad \ge q^{3/2}\left(\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_j)}}{\tau_j} + \log(M)\frac{\sigma^2}{\tau_j} + \varphi^2(\tau_j)\right),
\end{aligned}$$

from which the proof can easily be completed. □

Next we shall discuss the choice of $\tau_{\min}$ and $\tau_{\max}$. First, the natural domain of definition of the smoothness function φ is $(0, \|T^*T\|]$, so that the choice $\tau_{\max} = \|T^*T\|$ is natural. In this case the size of

$$\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau_{\max})$$

is at least $\varphi^2(\|T^*T\|)$ no matter how small the noise level σ was. The next result indicates that we can find $\tau_{\min}$ in such a way that we can remove the restriction to $\tau > \tau_{\min}$ if there is some "minimal" smoothness in the alternative.

Lemma 6.3. *Let $\tau_{\min} = \tau_{\min}(M)$ satisfy*

$$\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} \ge 1. \tag{6.4}$$

If the smoothness φ obeys $\varphi(\tau_{\min}) \le 1$ then for $0 < \tau \le \tau_{\min}$ we have

$$\begin{aligned}
&\log(M)\sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau) \\
&\quad \ge \frac{1}{2}\left(\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} + \log(M)\frac{\sigma^2}{\tau_{\min}} + \varphi^2(\tau_{\min})\right).
\end{aligned}$$

Proof. For $\tau < \tau_{\min}$ this easily follows from

$$\begin{aligned}
\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \varphi^2(\tau) &\ge \sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} \ge 1 \\
&\ge \frac{1}{2}\left(\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} + \varphi^2(\tau_{\min})\right),
\end{aligned}$$

which proves the assertion. □

Remark 6.2. For given $\sigma > 0$ the condition from (6.4) can always be satisfied. Below we shall further specify this as follows. If $\tau_{\max}$ is chosen as $\|T^*T\|$ then $\mathcal{N}(\tau_{\max}) \geq 1/2$, so that

$$\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} \geq \frac{\sqrt{\log(M)}\sigma^2}{\sqrt{2}\tau_{\min}} = \frac{\sqrt{\log(M)}\sigma^2 q^{1-M}}{\sqrt{2}\,\|T^*T\|} \geq \frac{1}{\sqrt{2}\,\|T^*T\|}\frac{\sigma^2}{q^M}.$$

Thus condition (6.4) holds for

$$M \geq \log_{1/q}\left(\sqrt{2}\,\|T^*T\|\,\right) + \log_{1/q}(1/\sigma^2).$$

We summarize the above considerations.

Proposition 6.2. *Suppose that M and $\tau_{\min}$ are chosen so that* (6.4) *holds. If the smoothness function φ obeys $\varphi(\tau_{\min}) \leq 1$ then*

$$\inf_{\tau\in\Delta_q}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right)$$
$$\leq q^{-3/2}2 \inf_{0<\tau\leq\tau_{\max}}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \log(M)\frac{\sigma^2}{\tau} + \varphi^2(\tau)\right).$$

The following result summarizes the above considerations; it asserts that the test $\Phi_\alpha^\star$ appears to be minimax (up to a log log term) in many cases.

Theorem 6.1. *Let α, β be fixed and $\Phi_\alpha^\star$ the test defined in* (6.3)*. Suppose that $\tau_{\max} = \|T^*T\|$, $\tau_{\min}$ is chosen so that $M \geq \log_{1/q}\left(\sqrt{2}\,\|T^*T\|\right) + \log_{1/q}(1/\sigma^2)$. Let τ_* be given from*

$$\varphi^2(\tau_*) = \sigma^2\sqrt{\log\log_{1/q}\left(\frac{1}{\sigma^2}\right)}\frac{\sqrt{\mathcal{N}(\tau_*)}}{\tau_*}. \tag{6.5}$$

If the underlying smoothness obeys $\varphi(\tau_{\min}) \leq 1$ and if

$$\frac{\log\log(\frac{1}{\sigma^2})}{\mathcal{N}(\tau_*)} = o(1) \qquad \text{as} \quad \sigma \to 0, \tag{6.6}$$

then there is a constant $C > 0$ such that

$$\rho^2(\Phi_\alpha^\star, \beta, \mathcal{E}_\varphi) \leq C \inf_{0<\tau\leq\tau_{\max}}\left(\sigma^2\sqrt{\log\log_{1/q}\left(\frac{1}{\sigma^2}\right)}\frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \varphi^2(\tau)\right).$$

In particular, as $\sigma \searrow 0$ we have $\tau_* \searrow 0$, and hence there is a constant $D = D(\alpha, \beta)$ such that

$$\rho(\Phi_\alpha^\star, \beta, \mathcal{E}_\varphi) \leq D\varphi(\tau_*) \quad \text{as } \sigma \searrow 0.$$

We shall indicate that assumption (6.6) is valid in many cases.

Lemma 6.4. *If there is a constant $c > 0$ such that the effective dimension obeys*

$$\mathcal{N}(\tau) \geq c\log(1/\tau), \tag{6.7}$$

and if the smoothness function φ increases at least as

$$\varphi(\tau) \leq \left(\log\log_{1/q}(1/\tau)\right)^4 \tag{6.8}$$

as $\tau \to 0$, then (6.6) *is valid.*

Proof. The parameter τ_* is determined from (6.5), and under (6.8) we find that

$$\sigma^4 \log\log_{1/q}\left(\frac{1}{\sigma^2}\right) = \frac{\tau_*^2 \varphi^4(\tau_*)}{\mathcal{N}(\tau_*)} \le \frac{\tau_*^2 \log\log_{1/q}(1/\tau_*)}{\mathcal{N}(\tau_*)} \le \tau_*^2 \log\log_{1/q}\left(\frac{1}{\tau_*}\right),$$

provided that τ_* is small enough. Monotonicity implies that $\sigma^2 \le \tau_*$. But then $\log\log(1/\sigma^2) \le \log\log(1/\tau_*)$, and we conclude that

$$\frac{\log\log(1/\sigma^2)}{\mathcal{N}(\tau_*)} \le \frac{\log\log(1/\tau_*)}{\mathcal{N}(\tau_*)} \le \frac{1}{c}\frac{\log\log(1/\tau_*)}{\log(1/\tau_*)} = o(1)$$

as σ, and hence τ_* tend to zero. □

Remark 6.3. This result covers many of the interesting cases, in particular the ones from Examples 4—5. In these cases Theorem 6.1 exhibits that the separation radii obey

$$\rho(\Phi_\alpha^*, \beta, \mathcal{E}_\varphi) \le D\Big(\sigma^2\sqrt{\log\log\frac{1}{\sigma^2}}\Big)^{s/(2s+2t+1/2)}, \quad \text{and}$$

$$\rho(\Phi_\alpha^*, \beta, \mathcal{E}_\varphi) \le D\log^{-s}(1/\sigma^2),$$

respectively. In particular, we see that adaptation does not demand an additional price for severely ill-posed testing problems.

Remark 6.4. A similar approach can be used when basing the adaptive test on a family of projection schemes. In this case we use a finite family of dimensions

$$\Delta_{2,j_0} := \{m = 2^{j+j_0},\ j = 0, \dots, M-1\},$$

and consider projection schemes with spaces $X_m,\ Y_{n(m)}$ for $m \in \Delta_{2,j_0}$. The above reasoning applies, taking into account the correspondence between regularization parameter τ in linear regularization schemes, and dimensions $m \sim 1/\tau$. For the sake of brevity, this will not be discussed in this paper.

APPENDIX A. PROOFS

A.1. Proof of Lemma 3.1

First, we propose an upper bound on $t_{R,\alpha}$. Notice that under H_0, $\|RY\|_H^2 = \|\sigma R\xi\|_H^2$. Then we get

$$\begin{aligned} P_{H_0}&\big(\|RY\|_H^2 - S_R^2 > 2\sqrt{2x_\alpha}S_R v_R + 2v_R^2 x_\alpha\big) \\ &= P_{H_0}\big(\|\sigma R\xi\|_H^2 - S_R^2 > 2\sqrt{2x_\alpha}S_R v_R + 2v_R^2 x_\alpha\big) \\ &= P_{H_0}\big(\|\sigma R\xi\|_H^2 - \mathbb{E}\|\sigma R\xi\|_H^2 > 2\sqrt{2x_\alpha}S_R v_R + 2v_R^2 x_\alpha\big) \\ &\le \exp\Big(-\frac{2x_\alpha v_R^2}{2v_R^2}\Big) = \alpha, \end{aligned}$$

where we have used Lemma B.1 (Appendix B) with $x = \sqrt{2x_\alpha}v_R$, in order to get the last inequality. Hence,

$$P_{H_0}\big(\|RY\|_H^2 - S_R^2 > 2\sqrt{2x_\alpha}S_R v_R + 2v_R^2 x_\alpha\big) \le \alpha,$$

which leads to the desired result.

Now, we turn our attention to the second term. We showed the relation of the problem to a specific sequence space model in (3.5). For this model we can apply Lemma B.2, which gives

$$P\big(\ \|RY\|_H^2 - \mathbb{E}\,\|RY\|_H^2 \le -2\sqrt{\Xi x_\beta}\ \big) \le \beta,$$

which completes the proof. □

A.2. Proof of Lemma 4.2

For the first assertion we bound, given an $\alpha > 0$ and using the singular numbers s_j of the operator T, the trace as follows. We abbreviate, for $s_j \geq \hat{c}\alpha$ the value $\beta_j := s_j/\hat{c}$. Then for any $0 < \underline{c} < 1$ we find that

$$\operatorname{tr}\left[\tau g_\tau^2(T^*T)T^*T\right] = \sum_{j=1}^{\infty} \tau g_\tau^2(s_j^2)s_j^2 \geq \sum_{\underline{c}s_j^2 \leq \hat{c}\tau \leq s_j^2} \tau g_\tau^2(s_j^2)s_j^2$$

$$\geq \sum_{\underline{c}\beta_j \leq \tau \leq \beta_j} \tau g_{\beta_j}^2(\hat{c}\beta_j)\hat{c}\beta_j \geq \frac{(\hat{\gamma})^2}{\hat{c}} \sum_{\underline{c}s_j^2 \leq \hat{c}\tau \leq s_j^2} \underline{c} \to \infty \quad \text{as } \tau \to 0.$$

Finally, by Lemma 4.1 we find that

$$\frac{v_\tau^2}{S_\tau^2} \leq \gamma_*^2 \frac{1}{\tau \operatorname{tr}\left[g_\tau^2(T^*T)T^*T\right]},$$

and the second assertion is a consequence of the first one. □

APPENDIX B. INEQUALITIES FOR GAUSSIAN ELEMENTS IN HILBERT SPACE

Lemma B.1. *Let X be a Gaussian random variable taking values in H. Then, for all $x > 0$,*

$$P\Big(\|X\|_H^2 - \mathbb{E}\,\|X\|_H^2 \geq x^2 + 2x\sqrt{\mathbb{E}\,\|X\|_H^2}\ \Big) \leq \exp\Big(-\frac{x^2}{2v^2}\Big),$$

where

$$v^2 := \sup_{\|\omega\|_H \leq 1} \mathbb{E}|\langle X, \omega\rangle|^2.$$

Proof. Using the Cauchy–Schwarz inequality, we first observe that

$$\big(\mathbb{E}\left[\|X\|_H\right] + x\big)^2 \leq \mathbb{E}\|X\|_H^2 + x^2 + 2x\sqrt{\mathbb{E}\|X\|_H^2}.$$

Hence, we get

$$P\Big(\|X\|_H^2 - \mathbb{E}\|X\|_H^2 \geq x^2 + 2x\sqrt{\mathbb{E}\,\|X\|_H^2}\ \Big)$$
$$\leq P\big(\|X\|_H^2 \geq (\mathbb{E}[\|X\|_H] + x)^2\big) = P\big(\|X\|_H \geq \mathbb{E}[\|X\|_H] + x\big) \leq \exp\Big(-\frac{x^2}{2v^2}\Big),$$

where for the last inequality we have used [19, Lemma 3.1]. □

Lemma B.2. *Let RY be as in* (2.2) *with expansion* (3.5). *Then*

$$P_f\big(\|RY\|_H^2 - \mathbb{E}_f\|RY\|_H^2 \leq -2\sqrt{\Xi x_\beta}\ \big) \leq \beta,$$

where Ξ is from (3.6).

Proof. The proof is a direct extension of the one proposed in [18] for a spectral cut-off approach. □

REFERENCES

1. Ya. Baraud, "Non-Asymptotic Minimax Rates of Testing in Signal Detection", Bernoulli **8** (5), 577–606 (2002).
2. Ya. Baraud, S. Huet, and B. Laurent, "Adaptive Tests of Linear Hypotheses by Model Selection", Ann. Statist. **31**, 225–251 (2003).
3. N. Bissantz, Th. Hohage, A. Munk, and F. H. Ruymgaart, "Convergence Rates of General Regularization Methods for Statistical Inverse Problems and Applications", SIAM J. Numer. Anal. **45** (6), 2610–2636 (2007).
4. G. Blanchard and P. Mathé, "Discrepancy Principle for Statistical Inverse Problems with Application to Conjugate Gradient Regularization", Inverse Problems **28** (11), 115011 (2012).
5. R. Bojanic and E. Seneta, "A Unified Theory of Regularly Varying Sequences", Math. Z. **134**, 91–106 (1973).
6. A. Caponnetto, *Optimal Rates for Regularization Operators in Learning Theory*, Techn. Rep. CSAIL-TR 2006-062 (Massachusetts Inst. of Technology, 2006).
7. L. Cavalier, "Inverse Problems in Statistics", in *Lect. Notes Statist. Proc.*, Vol. 203: *Inverse Problems and High-Dimensional Estimation* (Springer, Heidelberg, 2011), pp. 3–96.
8. C. de Boor, "Bounding the Error in Spline Interpolation", SIAM Rev. **16**, 531–544 (1974).
9. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, in *Mathematics and its Applications* (Kluwer Academic Publishers Group, Dordrecht, 1996), Vol. 375.
10. M. Fromont and B. Laurent, "Adaptive Goodness-of-Fit Tests in a Density Model", Ann. Statist. **34**, 1–45 (2006).
11. B. Hofmann and P. Mathé, "Analysis of Profile Functions for General Linear Regularization Methods", SIAM J. Numer. Anal. **45** (3), 1122–1141 (electronic) (2007).
12. Yu. I. Ingster, "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives. I", Math. Methods Statist. **2** (2), 85–114 (1993).
13. Yu. I. Ingster, Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives. II", Math. Methods Statist. **2** (3), 171–189 (1993).
14. Yu. I. Ingster, "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives. III", Math. Methods Statist. **2** (4), 249–268 (1993).
15. Yu. I. Ingster, T. Sapatinas, and I. A. Suslina, "Minimax Signal Detection in Ill-Posed Inverse Problems", Ann. Statist. **40**, 1524–1549 (2012).
16. Q. Jin and P. Mathé, "Oracle Inequality for a Statistical Raus–Gfrerer-Type Rule", SIAM/ASA J. Uncertainty Quantification **1** (1), 386–407 (2013).
17. B. Laurent, J.-M. Loubes, and C. Marteau, "Testing Inverse Problems: A Direct or an Indirect Problem?", J. Statist. Plann. Inference **141** (5), 1849–1861 (2011).
18. B. Laurent, J.-M. Loubes, and C. Marteau, "Non-Asymptotic Minimax Rates of Testing in Signal Detection with Heterogeneous Variances", Electron. J. Statist. **6**, 91–122 (2012).
19. M. Ledoux and M. Talagrand, *Probability in Banach Spaces. Isoperimetry and Processes* (Springer-Verlag, Berlin, 1991).
20. P. Mathé and B. Hofmann, "How General are General Source Conditions?", Inverse Problems **24** (1), 015009, 5 (2008).
21. P. Mathé and N. Schöne, "Regularization by Projection in Variable Hilbert Scales", Appl. Anal. **87** (2), 201–219 (2008).
22. P. Mathé and U. Tautenhahn, "Interpolation in Variable Hilbert Scales with Application to Inverse Problems", Inverse Problems **22** (6), 2271–2297 (2006).
23. F. Natterer, "Regularisierung schlecht gestellter Probleme durch Projektionsverfahren", Numer. Math. **28** (3), 329–341 (1977).
24. V. Spokoiny, "Adaptive Hypothesis Testing Using Wavelets", Ann. Statist. **24**, 2477–2498 (1996).
25. G. M. Vaĭnikko and U. A. Khyamarik. "Projection Methods and Self-Regularization in Ill-Posed Problems", Izv. Vyssh. Uchebn. Zaved. Mat. **84** (10), 3–17 (1985).
26. Tong Zhang. "Learning Bounds for Kernel Regression Using Effective Data Dimensionality", Neural Comput. **17** (9), 2077–2098 (2005).