HAL

archives-ouvertes.fr

General regularization schemes for signal detection in inverse problems

Clément Marteau, Peter Mathé

▶ To cite this version:

Clément Marteau, Peter Mathé. General regularization schemes for signal detection in inverse problems. 2013. <hal-00807335>

HAL Id: hal-00807335

https://hal.archives-ouvertes.fr/hal-00807335

Submitted on 3 Apr 2013

GENERAL REGULARIZATION SCHEMES FOR SIGNAL DETECTION IN INVERSE PROBLEMS

CLÉMENT MARTEAU AND PETER MATHÉ

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 (up to a log log factor) a test which adapts to the unknown smoothness in the alternative. Moreover, the authors discuss how the so-called *direct* and *indirect* tests are related via interpolation properties.

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. Precisely, we consider a statistical problem in Hilbert space, where we are given two Hilbert spaces H and K along with a (compact) linear operator $T: H \to K$. Given the (unknown) element $f \in H$ we observe

$$(1.1) \qquad Y = Tf + \sigma\xi,$$

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 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 $\|\xi\| \leq 1$). 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

$$\langle Y, g \rangle = \langle Tf, g \rangle + \sigma\langle \xi, g \rangle,$$

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.

In this study, our aim is to test the null hypothesis that the (underlying true) signal f corresponds to a given signal f_0 against a non-parametric alternative. More formally, we test

$$(1.2) \qquad H_0: f = f_0, \text{ against } H_{1,\rho}: \ f - f_0 \in \mathcal{E}, \ \|f - f_0\| \geq \rho,$$

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. Following the setting, (1.2) is known as a goodness-of-fit or a signal detection (when $f_0 = 0$)

testing problem. In the direct case, i.e. when $T = Id$, this problem has been widely investigated. We mention for instance seminal investigations proposed in [12, 13, 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 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 the same truncated singular value decomposition (see Section 3.1 for more details). Here we shall consider general *linear estimators* $\hat{f} = RY$, using the data Y. Plainly, since f_0 and hence Tf_0 are given, we can constrain the analysis to testing whether $f = 0$ (no signal) against the alternative $H_{1,\rho} : f \in \mathcal{E}, \|f\| \geq \rho$, 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 (1.2)) for which a prescribed level for the second kind error can be attained. More formally, given a fixed value of $\beta \in]0, 1[$ and a level-α test, we are interested in the radius $\rho(\Phi_\alpha, \beta, \mathcal{E})$ defined as

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

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.

Lower (and upper) bounds have already been established in order to characterize the behavior of this radius for different kind of smoothness assumptions (see for instance [15] or [18]). Recent analysis of (classical) inverse problems adopts a different approach by measuring the smoothness inherent in the class $\mathcal{E}$ *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 § 3.

Also, previous analysis was restricted to the truncated singular value decomposition of the underlying operator T. This limits the applicability of the test procedures, since often a singular value decomposition is hardly available, for instance when considering partial differential equations on domains with noisy boundary data. Therefore, the objective in this study is to propose alternative testing procedures that match the previous minimax bounds.

To this end we first consider general linear regularization in terms of an operator R (Sections 2.2 & 2.3), and we shall then specify these as linear regularization (in Section 3.1) or projection schemes (in Section 3.2), respectively. In each case, we derive the corresponding minimax separation radii. Next the relation between testing based on the estimation of f (inverse test), and test based on the estimation of Tf (direct test) is discussed in Section 4. 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 in Section 5 an adaptive test, which is based on a finite family of non-adaptive tests. It will be shown that this adaptive test, with an appropriately constructed finite family, is (up to a $\log \log$ factor) as good as the best among the whole family of tests.

2. Construction and calibration of the test

Considering the testing problem (1.2), most of the related tests are based on an estimation of $\|f\|^2$ ($\|f - f_0\|^2$ in the general case). Then, the idea is to reject H_0 as soon as this estimation becomes too large with respect to a prescribed threshold. As outlined above, in order to estimate $\|f\|^2$ where $f \in H$, from the observations Y, cf. (1.1), we shall use a general linear reconstruction operator $R \colon K \to H$.

2.1. **Notation and assumptions.** First we will specify the assumption on the noise ξ in (1.1).

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

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

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. Under such assumption, 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 3.1 & 3.2, we shall always make sure that this is the case. Then the application of R to the data Y may be decomposed as

$$(2.1) \qquad RY = RTf + \sigma R\xi = f_R + \sigma R\xi, \quad f \in H,$$

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

$$(2.2) \qquad \mathbb{E} \|RY\|^2 = \|RTf\|^2 + \sigma^2 \mathbb{E} \|R\xi\|^2 = \|f_R\|^2 + S_R^2,$$

where we introduce the *variance* of the estimator as

$$(2.3) \qquad S_R^2 := \sigma^2 \mathbb{E} \|R\xi\|^2 = \sigma^2 \operatorname{tr}\left[R^*R\right],$$

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

$$(2.4) \qquad v_R^2 := \sigma^2 \sup_{\|w\| \leq 1} \mathbb{E} |\langle R\xi, w \rangle|^2 = \sigma^2 \|R\|^2.$$

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 we have given R in terms of its singular value decomposition as

$$(2.5) \qquad Rg = \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, g \rangle \phi_j, \quad g \in K,$$

where we assume that both sequences $\{\psi_j\}_{j \in \mathbb{N}}$ and $\{\phi_j\}_{j \in \mathbb{N}}$ are orthonormal bases in K and H, respectively. Moreover, the sequence λ_j, $j = 1, 2, \ldots$ is assumed non-negative and arranged in non-increasing order. Then the following is well-known.

Lemma 2.1. *Let R be as in (2.5). 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 $v_R^2 \le S_R^2$, and typically these quantities differ by order. Some explicit computations will be provided below.

2.2. Construction of the test and control of the first kind error.

We see from (2.2) that the quantity $\|RY\|^2 - S_R^2$ is an unbiased estimator for the norm of $\|f_R\|^2$. If R is chosen appropriately, this term is an approximation of $\|f\|^2$, whose value is of first importance when considering the problem (1.2). Therefore, we shall use a threshold for $\|RY\|^2 - S_R^2$ to describe the test.

Let $\alpha \in (0,1)$ be the prescribed level for the first kind error, and we agree to abbreviate $x_\alpha := \log(1/\alpha)$. We define the test $\Phi_{\alpha,R}$ as

$$(2.6) \qquad\qquad \Phi_{\alpha,R} = \mathbf{1}_{\left\{\|RY\|^2 - S_R^2 > t_{R,\alpha}\right\}},$$

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

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

We emphasize that under H_0 the distribution of $\|RY\|^2 - S_R^2 = \sigma^2(\|R\xi\|^2 - \operatorname{tr}[R^*R])$ only depends on the chosen reconstruction R. Hence the quantile can be determined, at least approximately. Proposition 2.1 below establishes an upper bound for this quantile.

Proposition 2.1. *Let α be a fixed level. Then*

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

where the quantities S_R^2 and v_R^2 have been introduced in (2.3) and (2.4).

Proof. First notice that under H_0, $\|RY\|^2 = \|\sigma R\xi\|^2$. Then we get

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

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

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

which leads to the desired result. $\qquad\qquad\square$

2.3. Controlling the second kind error.

Here, our aim is to control the second kind error by some prescribed level $\beta > 0$, and again we abbreviate $x_\beta := \log(1/\beta)$. To this end, we have to exhibit conditions on f for which the probability $P_f(\Phi_\alpha = 0)$ will be bounded by β. By construction of the above test this amounts to bounding

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

where the latter follows from (2.2). In this section, we will investigate the lowest possible value of $\|f_R\|^2$ for which the previous 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 $\|RY\|^2 - S^2$. In other words

$$(2.8) \qquad P_f\left(\|RY\|^2 - \mathbb{E}\|RY\|^2 \leq t_{R,\beta}(f)\right) = \beta.$$

Then, we get from (2.7) and (2.8) that $P_f(\Phi_{\alpha,R} = 0)$ will be bounded by β as soon as

$$(2.9) \qquad t_{R,\alpha} - \|f_R\|^2 \leq t_{R,\beta}(f) \Leftrightarrow \|f_R\|^2 \geq t_{R,\alpha} - t_{R,\beta}(f).$$

We have already an upper bound on the $1 - \alpha$-quantile $t_{\alpha,R}$. In order to conclude this discussion, we need a lower bound on $t_{R,\beta}(f)$.

Lemma 2.2. *Let the reconstruction R be given as in (2.5), and let*

$$(2.10) \qquad \Sigma := \sum_{j=1}^{\infty} \sigma_j^4 + 2\sum_{j=1}^{\infty} \sigma_j^2 \theta_j^2.$$

Then

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

Proof. We first show the relation of the problem to a specific sequence space model. By construction of R, using (2.5), we can expand

$$RY = \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, Y \rangle \phi_j \;\; = \;\; \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, Tf \rangle \phi_j + \sigma \sum_{j=1}^{\infty} \lambda_j \langle \psi_j, \xi \rangle \phi_j,$$

$$= \;\; \sum_{j=1}^{\infty} \theta_j \phi_j + \sum_{j=1}^{\infty} \sigma_j \varepsilon_j \phi_j,$$

where $\theta_j := \lambda_j \langle \psi_j, Tf \rangle$ and $\sigma_j := \sigma \lambda_j$ for all $j \in \mathbb{N}$, and the ε_j are i.i.d. standard Gaussian random variables. Then, we can apply Lemma A.2, which gives

$$P(\|RY\|^2 - \mathbb{E}\|RY\|^2 \leq -2\sqrt{\Sigma x_\beta}) \leq \beta,$$

which completed the proof. $\qquad\square$

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

$$(2.11) \qquad C_{\alpha,\beta}^* = (4\sqrt{x_\beta} + 4\sqrt{2x_\alpha}),$$

which is a function of α and β, only.

Proposition 2.2. *Let us consider the test $\Phi_{\alpha,R}$ as introduced in (2.6), and let*

$$(2.12) \qquad r^2(\Phi_{\alpha,R}, \beta) := C_{\alpha,\beta}^* Sv + (4x_\alpha + 8x_\beta)v^2.$$

Then

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

Proof. The equation (2.9) provides a condition for which $P_f(\Phi_\alpha = 0) \leq \beta$. Using Proposition 2.1 and Lemma 2.2, we see that this condition is satisfied as soon as

$$\|f_R\|^2 \geq 2\sqrt{\Sigma x_\beta} + 2\sqrt{2x_\alpha}Sv + 2v^2 x_\alpha.$$

Now we bound

$$\Sigma \;\; = \;\; \sigma^4 \sum_{j=1}^{+\infty} \lambda_j^4 + 2\sigma^2 \sum_{j=1}^{+\infty} \lambda_j^2 \times \lambda_j^2 \langle \psi_j, Tf \rangle^2,$$

$$\leq \;\; S^2 v^2 + 2v^2 \|f_R\|^2.$$

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

$$
\begin{aligned}
2\sqrt{\Sigma x_\beta} &\leq 2Sv\sqrt{x_\beta} + 2\sqrt{2x_\beta}\|f\|v, \\
&\leq 2Sv\sqrt{x_\beta} + \frac{1}{2}\|f\|^2 + 4x_\beta v^2.
\end{aligned}
$$

In particular, the condition (2.9) will be satisfied as soon as

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

$\square$

Remark 2.1. Please note that the condition on $\|f_R\|^2$ is (as most of the results presented below) non-asymptotic, i.e. we do not require that $\sigma \to 0$. Using, the property $v \leq S$, we can obtain the simple bound

$$
(2.13) \qquad r^2(\Phi_\alpha, \beta) \leq C_{\alpha,\beta} Sv, \text{ where } C_{\alpha,\beta} = 4\sqrt{x_\beta} + 4\sqrt{2x_\alpha} + 4x_\alpha + 8x_\beta.
$$

In an asymptotic setting, the value of the constant $C_{\alpha,\beta}$ may sometimes be improved. In particular, the majorization $v \leq S$ is rather rough. In many cases, we will only deal with the constant $C^*_{\alpha,\beta}$, and we refer to Corollary 3.1.

3. Determining the separation radius under smoothness

We have seen in the previous section that we need to have that $\|f_R\|^2 \geq C_{\alpha,\beta}Sv$ in order to control the second kind error. Nevertheless, the alternative in (1.2) is expressed in term of a lower bound on $\|f\|^2$. In this section, we take advantage on the smoothness of f in order to propose a upper bound on the separation radius.

Using a triangle inequality, we obtain

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

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

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

In other words, we get from Proposition 2.2 that

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

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

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

where the reconstructions R belong to certain families $\mathcal{R}$. We shall establish order optimal bounds in two cases, the case of linear regularization and by using projection schemes.

As 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 for more general (operator) functions $\varphi(T^*T)$. The corresponding operator $\varphi(T^*T)$ is compact, whenever, $\varphi(t) \to 0$ as $t \to 0$. Therefore, we shall restrict to functions with this property. Precisely, we let

$$
(3.3) \qquad \mathcal{E}_\varphi = \{h \in H, \ h = \varphi(T^*T)\omega, \text{ for some } \|\omega\| \leq 1\},
$$

for a continuous non-decreasing function φ which obeys $\varphi(0) = 0$ (index function), be a *general source set*. It was established in [20] that each element in H has some

smoothness of this kind, and hence the present approach is most general. Examples, which relate Sobolev type balls to the present setup are given in Examples 3 & 4.

3.1. **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 can be tuned in order to obtain suitable tests.

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

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

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

(1) *For each $0 < t \le \|T^*T\|$ we have that $|r_\tau(t)| \to 0$ as $\tau \to 0$;*
(2) *There is a constant γ_1 such that $\sup_{0 \le t \le \|T^*T\|} |r_\tau(t)| \le \gamma_1$ for all $0 < \tau \le \|T^*T\|$;*
(3) *There is a constant $\gamma_* \ge 1$ such that $\sup_{0 \le t \le \|T^*T\|} \tau |g_\tau(t)| \le \gamma_*$ for all $0 < \tau < \infty$,*

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

Notice, that in contrast to the usual convention we used the symbol τ instead of α, as the latter is used as control parameter for the error of the first kind.

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 $f_R = f_\tau = g_\tau(T^*T)T^*Tf$.

Example 1 (truncated svd, spectral cut-off). Let $(s_j, u_j, v_j)_{j \in \mathbb{N}}$ be the singular value decomposition of the operator T, i.e., we have that

$$Tf = \sum_{j=1}^\infty s_j \langle f, u_j \rangle v_j, \quad f \in H,$$

and the singular numbers $s_1 \ge s_2 \cdots \ge 0$ are arranged in decreasing order. With this notation we can use the function $g_\tau(t) := 1/t$, $t \ge \tau$ and zero else. This means that we approximate the inverse mapping of T by the finite expansion $R_\tau Y := \sum_{s_j^2 \ge \tau} \frac{1}{s_j} \langle Y, v_j \rangle u_j$, $Y \in K$, The condition $s_j^2 \ge \tau$ translates to an upper bound $1 \le j \le D = D(\tau)$. The element f_τ is then given as $f_\tau = \sum_{j=1}^D \langle f, u_j \rangle u_j$.

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

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 (2.3), (2.4) (with a slight abuse of notation). To this end, we will impose the following assumption.

Assumption A2. The operator T is a Hilbert-Schmidt operator, i.e.,

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

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 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*

(3.4) $$\mathcal{N}(\lambda) := \mathrm{tr}\,\left[(T^*T + \lambda I)^{-1} T^*T\right]$$

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

By Assumption A2 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].

$$(3.5) \qquad \operatorname{tr}\left[g_\tau^2(T^*T)T^*T\right] \leq 2\gamma_*^2 \frac{\mathcal{N}(\tau)}{\tau},$$

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

Lemma 3.1. *Let $R_\tau := g_\tau(T^*T)T^* : K \to H$. Assume that Assumption A2 holds, then we have that*

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

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

Proof. The proof is a direct consequence of the definition of S_τ^2, v_τ^2 and of (3.5). $\square$

The previous lemma only provides upper bounds for the terms 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.

Lemma 3.2. *Suppose that the regularization g_τ has the following properties.*

(1) *There are constants $\hat{c}, \hat{\gamma} > 0$ such that $|g_\tau(\hat{c}\tau)| \geq \hat{\gamma}/\tau$ for $\alpha > 0$, and*
(2) *for each $0 < t \leq \|T^*T\|$ the function $\tau \to |g_\tau(t)|$ is decreasing.*

If the singular numbers of the operator T decay moderately, such that $\#\left\{j,\ \hat{c}\tau \leq s_j^2 \leq \hat{c}/\underline{c}\tau\right\} \to \infty$ 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 that $v_\tau/S_\tau \to 0$ as $\tau \to 0$.

Proof. 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 3.1 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. $\square$

Remark 3.1. The assumptions which are 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 hold for (at most) polynomial decay.

Lemma 3.2 implies that in some specified cases the separation radius defined in (2.12) is of size $C_{\alpha,\beta}^\star S_\tau v_\tau$ as $\tau \to 0$. This is summarized in the following corollary.

Corollary 3.1. *Let $C_{\alpha,\beta}^\star$ and $r^2(\Phi_{\alpha,R}, \beta)$ be as in (2.11) and (2.12), respectively. Under the assumptions of Lemma 3.2 we have that*

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

We turn to bounding the bias $\|f - f_\tau\|$. This can be done under the assumption that the chosen regularization has enough qualification, see e.g. [11].

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

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

Remark 3.2. 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, truncated svd has arbitrary qualification with constant $\gamma = 1$.

In this case we can bound the bias at $f_R = f_\tau$.

Proposition 3.1. *Let g_τ be any regularization having qualification φ with constant γ. If $f \in \mathcal{E}_\varphi$ then*

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

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

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

$\square$

Now we have established bounds for all quantities occurring in (3.2), and this yields the main result for linear regularization.

Theorem 3.1. *Assume that Assumption A2 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*

$$(3.6) \qquad \qquad \varphi^2(\tau) = \sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau}.$$

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

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

where the constant $C_{\alpha,\beta}^*$ has been introduced in (2.11). In particular, we get that

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

Proof. By Proposition 3.1 and Proposition 2.2, we have that

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

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

Remark 3.3. 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 (3.6) satisfies $\tau_* \to$ as $\sigma \to 0$. Since the effective dimension tends to infinity as $\tau \to 0$, we get that

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

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 [17]).

Example 3 (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 A2 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

$$(3.7) \qquad \mathcal{E}^{\mathcal{X}}_{a,2} := \left\{ f, \ \sum_{j=1}^\infty a_j^2 \langle f, \phi_j^2 \rangle \leq R^2 \right\}, \ \text{with } a_j = j^s, \quad \forall j > 1,$$

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 (3.6) is computed as $\tau_* \asymp \sigma^{8t/(4s+4t+1)}$, which results in an asymptotic separation rate of

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

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

Example 4 (severely ill-posed problem). Here we assume a decay of the form $s_k \asymp \exp(-\gamma k)$, $k \in \mathbb{N}$ of the singular numbers. The effective dimension behaves like $\mathcal{N}(\tau) \asymp \frac{1}{\gamma} \log(1/\tau)$. The Sobolev ball from (3.7) 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

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

again recovering the corresponding result from [18].

3.2. **Projection schemes.** Here we follow the ideas from the previous section. 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*. Throughout we shall denote 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

$$(3.8) \qquad Q_n Y = Q_n T P_m x + \sigma Q_n \xi.$$

Without further assumptions, the finite dimensional equation (3.8) 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

$$(3.9) \qquad f_{m,n} := (Q_n T P_m)^\dagger Q_n Y.$$

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 $m \to n(m)$, $m = 1, 2, \ldots$,*

then the corresponding sequence of mappings

$$(3.10) \qquad Y \to f_{m,n(m)} := (Q_n T P_m)^\dagger Y$$

is called projection scheme.

Example 5 (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 $H_m := \text{span} \{u_1, \ldots, u_m\} \subset H$, and $K_m := \text{span} \{v_1, \ldots, v_m\} \subset K$, respectively. In this case we see that $(Q_n T P_m)^\dagger Y = \sum_{j=1}^m \frac{1}{s_j} \langle Y, v_j \rangle$.

Henceforth we shall always assume that the mapping $(Q_n T P_m)^\dagger : K_n \to H_m$ is invertible, i.e., the related linear system of equations has a unique solution. This gives an (implicit) relation $n = n(m)$, typically $n = n$ will do. However, our subsequent analysis will be done using the dimension m of the space H_m for quantification. 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 need to control $\text{tr}[R^* R]$ as well as $\|R\|$. The latter is related to the robustness (stability) of the 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*

$$(3.11) \qquad \left\|(Q_n T P_m)^\dagger\right\| \leq \frac{D_R}{j(T, H_m)}, \quad m = 1, 2, \ldots$$

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

$$(3.12) \qquad j(T, H_m) := \inf_{0 \neq x \in H_m} \frac{\|Tx\|}{\|x\|}.$$

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 $0 < C_B \leq 1$ such that

$$C_B s_m(T) \leq j(T, H_m).$$

We summarize our previous outline as follows.

Lemma 3.3. *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*

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

In particular we have that

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

We turn to bounding S_R^2. Before doing so we mention that for spectral cut-off from Example 5, this bound can easily be established.

Lemma 3.4. *For spectral cut-off we have*

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

In order to achieve 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 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 $-r$, for some $r \geq 0$, and we refer to [5] for a treatment. In particular this covers moderately ill-posed problems where $s_j \asymp j^{-r}$. We will not use the index r. However, if the sequence $s_j, j = 1, 2, \dots$, is regularly varying with index $-r$ then the sequence $s_j^{-2}, j = 1, 2, \dots$, is regularly varying with index $2r$, and we have that

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

In particular there is a constant C_r such that

$$(3.13) \qquad \frac{m}{s_m^2} \leq C_r^2 \sum_{j=1}^{m} \frac{1}{s_j^2},$$

and the latter bound is actually all that is needed.

Lemma 3.5. *Suppose that the sequence $s_j, j = 1, 2, \dots$, is such that for the constant C_r the estimate (3.13) holds. 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*

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

If, in addition the Assumption A2 is satisfied, then we have that

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

Proof. We notice that the mapping $\left((Q_n T P_m)^\dagger\right)^*$ 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

$$\operatorname{tr}\left[\left((Q_n T P_m)^\dagger\right)^* (Q_n T P_m)^\dagger\right] = \operatorname{tr}\left[(Q_n T P_m)^\dagger \left((Q_n T P_m)^\dagger\right)^*\right]$$

$$= \sum_{j=1}^{\infty} \left\|\left((Q_n T P_m)^\dagger\right)^* u_j\right\|^2$$

$$= \sum_{j=1}^{m} \left\|\left((Q_n T P_m)^\dagger\right)^* u_j\right\|^2$$

$$\leq m \left\|\left((Q_n T P_m)^\dagger\right)^*\right\|^2 = m \left\|(Q_n T P_m)^\dagger\right\|^2.$$

Using Lemma 3.3 we see that $\operatorname{tr}\left[\left((Q_n T P_m)^\dagger\right)^* (Q_n T P_m)^\dagger\right] \leq m \frac{D_R^2}{C_B^2} \frac{1}{s_m^2}$. Now we use (3.13) to complete the proof of the first assertion. Under Assumption A2 we continue and use the inequality $u/v \leq 2v/(u+v)$, $0 < u \leq v$, to see that

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

and the proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 3.4. Notice that Lemma 3.5 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 3.1 (with $\tau \leftarrow s_m^2$).

Next, we need to bound $\|f - f_R\|$, as this was done in § 3.1 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

$$(3.14) \qquad \left\| f - (Q_n T P_m)^\dagger T f \right\| \leq \left\| (Q_n T P_m)^\dagger T \right\| \|f - P_m f\|.$$

Therefore, we can bound the bias whenever the norms $\left\| (Q_n T P_m)^\dagger T \right\|$ 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 $\left\| (Q_n T P_m)^\dagger T \right\| \leq D_Q$.*

We emphasize that under quasi-optimality the bound for the bias entirely depends on the quality of the projections P_m with respect to the element f.

Definition 7 (Degree of approximation). *Suppose that $\{H_m\}$, $\dim(H_m) \leq m$, 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*

$$(3.15) \qquad \|(I - P_m)\varphi(T^*T)\| \leq C_D \varphi(s_{m+1}), \quad m = 1, 2, \dots$$

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, $m = 1, 2, \dots$, we refer to [21] for more details on degree of approximation and Bernstein-type bounds. We now can state the analogue of Proposition 3.1 for projection schemes.

Proposition 3.2. *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 that*

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

We now return to the problem raised in (3.2). 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 A2 is satisfied, i.e. that T is a Hilbert-Schmidt operator. If it is not the case, Theorem 3.2 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 3.2. *Suppose that the approximate solutions are obtained by a projection scheme which is quasi-optimal and robust and that Assumption A2 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*

$$(3.16) \qquad m_* = \max\left\{ m, \quad \varphi^2(s_m^2) \geq \sigma^2 \frac{\sqrt{\mathcal{N}(s_m^2)}}{s_m^2} \right\}.$$

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

$$\inf_m \left\{ r^2(\Phi_\alpha, \beta) + \|f - f_m\|^2 \right\}$$

$$\leq \left(C_{\alpha,\beta}^* \frac{D_R^2}{C_B^2} C_r + (4x_\alpha + 8x_\beta) \frac{D_R^2}{C_B^2} \frac{1}{\sqrt{\mathcal{N}(s_{m_*}^2)}} + D_Q^2 C_D^2 \right) \varphi^2(s_{m_*}^2),$$

where the constant $C^*_{\alpha,\beta}$ has been introduced in (2.11).

Proof. By using Lemma 3.5 and Proposition 3.2 we see that for any choice of discretization level m we have

$$r^2(\Phi_\alpha, \beta) + \|f - f_m\|^2$$

$$\leq C^*_{\alpha,\beta} \frac{D_R^2}{C_B^2} C_r \sigma^2 \frac{\sqrt{\mathcal{N}(s_m^2)}}{s_m^2} + (4x_\alpha + 8x_\beta)\sigma^2 \frac{D_R^2}{C_B^2} \frac{1}{s_m^2} + D_Q^2 C_D^2 \varphi^2(s_{m+1}^2)$$

$$\leq \left(C^*_{\alpha,\beta} \frac{D_R^2}{C_B^2} C_r + (4x_\alpha + 8x_\beta)\frac{D_R^2}{C_B^2} \frac{1}{\sqrt{\mathcal{N}(s_{m_*}^2)}} + D_Q^2 C_D^2 \right) \times$$

$$\times \max\left\{ \sigma^2 \frac{\sqrt{\mathcal{N}(s_m^2)}}{s_m^2}, \varphi^2(s_{m+1}^2) \right\}.$$

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

$$\varphi^2(s_{m_*+1}^2) \leq \varphi^2(s_{m_*}^2).$$

Also, by the choice of m_* we see that

$$\sigma^2 \frac{\sqrt{\mathcal{N}(s_{m_*}^2)}}{s_{m_*}^2} \leq \varphi^2(s_{m_*}^2),$$

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

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

Corollary 3.2. *Under the assumptions of Theorem 3.2 we get that*

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

as $\sigma \to 0$.

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

3.3. **Discussion.** We first highlight the important fact that in both cases (provided Assumption A2 is satisfied), linear regularization and for projection schemes the upper bound is obtained by solving the same 'equation', $\sigma^2 = \tau\varphi^2(\tau)/\sqrt{\mathcal{N}(\tau)}$, such that relating $\tau_* \sim s_{m_*}^2$, see Theorems 3.1 and 3.2. However, this function is different from the one used for *function estimation* in inverse problems. In the same setting the 'optimal' parameter τ_{est} is there obtained from solving

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

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

Another remark may be of interest. For the estimation problem, within the same context, the bias variance decomposition leads to a variance term S_R^2, and in order to achieve optimal order reconstruction, this will be calibrated with the function φ^2. As we have seen above, for testing the same calibration is done between the functions Sv and φ^2. Since, as already mentioned $Sv \leq S^2$ this calibration always yields a smaller value, which again explains the different rates for separation radius and estimation error.

Previous analysis of the spectral cut-off regularization scheme for testing in inverse problems revealed the importance of the quantity

$$(3.17) \qquad \rho_D := \left(\sum_{j=1}^{D} \frac{1}{s_j^4} \right)^{1/4}, \quad D = 1, 2, \dots$$

We mention the non-asymptotic lower and upper bounds, slightly adapted to the present setup, given for instance in [18] as

$$\rho^2(\mathcal{E}_\varphi, \alpha, \beta) \geq \sup_D \min \left\{ c_{\alpha,\beta}^2 \rho_D^2, \varphi^2(s_D^2) \right\},$$

$$\rho^2(\mathcal{E}_\varphi, \alpha, \beta) \leq \inf_D \left(C_{\alpha,\beta}^2 \rho_D^2 + \varphi^2(s_D^2) \right)$$

Thus the bounds established in this study are sharp whenever $\rho_D^2 \asymp S_D v_D$, where $S_D^2 = \sum_{j=1}^{D} s_j^{-2}$, and $v_D^2 = s_D^{-2}$, respectively. More explicitly, if

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

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 $s_j \asymp \exp(-\gamma j), j = 1, 2, \dots$, thus covering severely ill-posed problems. Remark that instead of the terms involved in (3.17), the quantities S_D and v_D have nice interpretation as strong and weak variances of the spectral cut-off schemes.

4. Relating the direct and inverse testing problem

For injective linear operators T, the assertions "$f = 0$" and "$Tf = 0$" are equivalent. Hence, testing $H_0 : f = 0$ or testing $H_0 : Tf = 0$ is 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

$$(4.1) \qquad H_0^I : f = 0, \text{ against } H_1^I : f \in \mathcal{E}_\varphi, \ \|f\|^2 \geq (\rho^I)^2,$$

while the direct testing problem corresponds to test

$$(4.2) \qquad H_0^D : Tf = 0, \text{ against } H_1^D : f \in \mathcal{E}_\varphi, \ \|Tf\|^2 \geq (\rho^D)^2.$$

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

4.1. **Relating the separation rates.** The authors in [17] discussed whether both problems are related. The main result, Theorem 1, ibid. asserts that for a variety of cases each minimax test Φ_α for the direct problem ($H_0 : Tf = 0$) is also minimax for the related inverse problem ($H_0 : f = 0$). This fundamental results is based in 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 4.1 (Interpolation inequality)**.** *Let φ be from (3.3), and let $\Theta(u) := \sqrt{u}\varphi(u), \ u > 0$. If the function $u \mapsto \varphi^2 \left(\left(\Theta^2 \right)^{-1} (u) \right)$ is concave then*

$$(4.3) \qquad \|f\| \leq \varphi \left(\Theta^{-1}(\|Tf\|) \right), \quad f \in \mathcal{E}_\varphi.$$

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

Theorem 4.1. *Let φ be an index function with related function Θ, such that the function $u \mapsto \varphi^2\left(\left(\Theta^2\right)^{-1}(u)\right)$ is concave. Let Φ_α be a level-α test for the direct problem $H_0^D : Tf = 0$ with uniform separation rate $\rho^D(\Phi_\alpha, \mathcal{E}_\Theta, \beta)$. Then Φ_α constitutes a level-α test for the inverse problem $H_0^I : f = 0$ with uniform separation rate*

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

Consequently we have for the minimax separation rates that

$$(4.4) \qquad \rho^I(\mathcal{E}_\varphi, \alpha, \beta) \leq \varphi\left(\Theta^{-1}\left(\rho^D(\mathcal{E}_\Theta, \alpha, \beta)\right)\right).$$

Proof. Clearly, the test Φ_α is a level-α test for both problems, and we need to control the second kind error. But if $\|f\| \geq \varphi\left(\Theta^{-1}\left(\rho^D(\Phi_\alpha, \mathcal{E}_\Theta, \beta)\right)\right)$ then Lemma 4.1 yields that $\|Tf\| \geq \rho^D(\Phi_\alpha, \mathcal{E}_\Theta, \beta)$, 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 $\varphi\left(\Theta^{-1}\left(\rho^D(\mathcal{E}_\Theta, \alpha, \beta)\right)\right)$, which gives (4.4). $\qquad\square$

Remark 4.1. In many cases the bound (4.4) actually is an asymptotic equivalence

$$(4.5) \qquad \varphi^{-1}\left(\rho^I(\mathcal{E}_\varphi, \alpha, \beta)\right) \asymp \Theta^{-1}\left(\rho^D(\mathcal{E}_\Theta, \alpha, \beta)\right), \quad \sigma \to 0.$$

It may be enlightening to see this on the base of Example 3. 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

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

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 $\rho(\mathcal{E}_\Theta, \alpha, \beta) \asymp \sigma^{2(s+t)/(2s+2t+1/2)}$, which in turn gives

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

and hence (4.5) for moderately ill-posed problems.

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

We emphasize that, by virtue of Theorem 4.1, any lower bound for the minimax separation rate in the inverse testing problem yields a lower bound for the corresponding direct problem.

Remark 4.2. Thanks to Theorem 4.1, it is possible to prove that in all the cases considered in this paper, a test minimax for (4.2) will be also minimax for (4.1). Nevertheless, the reverse is not true. We will not dwell into details, instead we refer to [17] for a detailed discussion on this subject.

4.2. **Designing tests for the direct problem.** The coincidence in (4.5) is not by chance and we indicate a further result in this direction. Recall from 3.1 that the value of $\tau_* = \tau_*^{\mathrm{IP}}$ was obtained from (3.6), and hence that we actually have $\rho(\mathcal{E}_\varphi, \alpha, \beta) \asymp \varphi(\tau_*^{\mathrm{IP}})$, such that the left hand side in (4.5) equals τ_*^{IP}. We shall see next that the corresponding value $\tau_* = \tau_*^{\mathrm{DP}}$ is obtained from the same equation (3.6) when basing the direct test on the family $\widehat{TR_\tau} = TR_\tau$ with family $R_\tau = g_\tau(T^*T)T^*$ as in § 3.1. Then $TR_\tau = g_\tau(TT^*)TT^*$, and we bound its variance and weak variance, next.

Lemma 4.2. *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 A2 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 $\|Tf - Tf_\tau\|$ with $f_\tau = g_\tau(T^*T)T^*Tf$

Lemma 4.3. *Assume that $f \in \mathcal{E}_\varphi$. If the regularization g_τ has qualification Θ with constant γ then*

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

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

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

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

We recall from § 3 the quantity $r^2(\Phi_\alpha, \beta) := C_{\alpha,\beta} S_\tau v_\tau$, where we now consider $\tilde{R}_\tau$ and $\tilde{v}_\tau$ from Lemma 4.2 for bounding $\|Tf\|^2 \geq C_{\alpha,\beta} \tilde{S}_\tau \tilde{v}_\tau$ from below.

Corollary 4.1. *Suppose that g_τ is a regularization which has qualification Θ, $f \in \mathcal{E}_\varphi$ and that Assumption A2 holds. Let τ_*^{DP} be chosen from the equation*

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

Then

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

We stress that the equation (4.6) for determining τ_*^{DP} is the same equation as (3.6), since $\Theta^2(\tau) = \tau\varphi^2(\tau)$, and this explains the identical asymptotics in (4.5) as being equal to $\tau_*^{DP} = \tau_*^{IP}$.

This result sheds light to 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 § 3.2 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 $S_m^2 = \text{tr}\,[Q_m^* Q_m] = m$, 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 $\|(I - Q_m)\Theta(TT^*)\| \leq C_D \Theta(s_{m+1}^2)$, $m = 1, 2 \dots$. 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

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

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

5. 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.

5.1. 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 A2. Then, we can associate to 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 $\Phi_\alpha^\star$ as

$$(5.1) \qquad \Phi_\alpha^\star = \max_{R \in \mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|}, R}.$$

The performance of $\Phi_\alpha^\star$ is summarized in the following proposition.

Proposition 5.1. *The test introduced in (5.1) is a level-α test. Moreover*

$$P_f(\Phi_\alpha^\star = 0) \leq \beta,$$

as soon as

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

where the term r^2 has been introduced in (2.12).

Proof. We first remark that

$$P_{H_0}(\Phi_\alpha^\star = 1) = P_{H_0}\left(\max_{R \in \mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|}, R} = 1 \right),$$

$$= P_{H_0}\left(\bigcup_{R \in \mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|}, R} = 1 \right),$$

$$\leq \sum_{R \in \mathcal{R}} P_{H_0}\left(\Phi_{\frac{\alpha}{|\mathcal{R}|}, R} = 1 \right) = \alpha,$$

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

$$P_f(\Phi_\alpha^\star = 0) = P_{H_0}\left(\max_{R \in \mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|}, R} = 0 \right),$$

$$= P_{H_0}\left(\bigcap_{R \in \mathcal{R}} \Phi_{\frac{\alpha}{|\mathcal{R}|}, R} = 0 \right),$$

$$\leq \inf_{R \in \mathcal{R}} P_{H_0}\left(\Phi_{\frac{\alpha}{|\mathcal{R}|}, R} = 0 \right).$$

We can conclude using (3.1). $\qquad \square$

Proposition 5.1 proves that the detection radius associated to the test defined in (5.1) 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 5.1. *Let $C^*_{\alpha,\beta}$ the term introduced in (2.11). If the family $\mathcal{R}$ of regularization schemes has cardinality $M := |\mathcal{R}| \geq 1$, then*

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

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

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

$$\begin{aligned}
C^*_{\alpha,\beta} &= 2\sqrt{x_\beta} + 2\sqrt{2x_{\alpha/M}}, \\
&= C^*_{\alpha,\beta} + 2\left(\sqrt{2x_\alpha + 2\log(M)} - \sqrt{2x_\alpha}\right). \\
&\leq C^*_{\alpha,\beta} + 2\sqrt{2\log(M)}.
\end{aligned}$$

The second assertion is trivial, because $\log(M) > 1$ for $M \geq 4$. $\qquad\square$

Therefore, the price to pay for using Φ^*_α is a term of order $\sqrt{\log(|\mathcal{R}|)}$, up to some condition on the behavior of the effective dimension (see Theorem 5.1 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 such was established Theorems 3.1 and 3.2.

In the following, we propose practical situations where such an adaptive scheme can be used. In particular, we propose families of regularizations 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 5.1. In the test (5.1), each regularization operator $R \in \mathcal{R}$ is associated to 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.

5.2. **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 to each function g_τ the operator R_τ and we deal with the family $\mathcal{R} = (R_\tau)_{\tau>0}$. In order to apply Proposition 5.1 we need to specify a finite subset $\mathcal{R} \subset (0,\infty)$ on which the test Φ^*_α will be based on. To this end we will use an exponential grid. Given an initial value $\tau_{\max}$, and a tuning parameter $0 < q < 1$ we consider the exponential grid

$$(5.2) \qquad \Delta_q := \left\{\tau = q^j \tau_{\max}, \quad j = 0, \dots, M-1\right\}, \quad \text{for some } M > 1.$$

Then we use the adaptive test

$$(5.3) \qquad \Phi^*_\alpha = \max_{\tau \in \Delta_q} \Phi_{\alpha/M,\tau}.$$

The result from Proposition 5.1 can be rephrased as follows. By virtue of Lemma 5.1, and using the bounds from Lemma 3.1 & Proposition 3.1, respectively, we find that the test Φ^*_α bounds the error of the second kind by β as soon as

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

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 that

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

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

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

Lemma 5.2 (cf. [16, Proof of Thm. 3.1]). *We have that*

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

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

Proof. For any τ with $\tau_{\min} < \tau \leq \tau_{\max}$ we find an index $1 \leq j \leq M$ for which $\tau_j < \tau \leq \tau_j/q$. 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

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

$$\geq \sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau_j/q)}}{\tau_j/q}+\log(M)\frac{\sigma^2}{\tau_j/q}+\varphi^2(\tau_j)$$

$$= \sqrt{\log(M)}\sigma^2\left(\frac{\tau_j}{q}\right)^{-3/2}\left(\frac{\tau_j}{q}\right)^{3/2}\frac{\sqrt{\mathcal{N}(\tau_j/q)}}{\tau_j/q}+q\log(M)\frac{\sigma^2}{\tau_j}+\varphi^2(\tau_j)$$

$$\geq \sqrt{\log(M)}\sigma^2\left(\frac{\tau_j}{q}\right)^{-3/2}\tau_j^{3/2}\frac{\sqrt{\mathcal{N}(\tau_j)}}{\tau_j}+q^{3/2}\log(M)\frac{\sigma^2}{\tau_j}+\varphi^2(\tau_j)$$

$$\geq q^{3/2}\left(\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau_j)}}{\tau_j}+\log(M)\frac{\sigma^2}{\tau_j}+\varphi^2(\tau_j)\right),$$

from which the proof can easily be completed. $\qquad\square$

We shall next discuss the choices of τ_{min} and $\tau_{\max}$. First, the natural domain of definition of the smoothness function φ is on $(0, \|T^*T\|]$, such 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 5.3. *Let* $\tau_{\min} = \tau_{\min}(M)$ *satisfy*

$$(5.4)\qquad\qquad\qquad\sqrt{\log(M)}\sigma^2\frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}}\geq 1.$$

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

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

$$\geq \frac{1}{2}\left(\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} + \log(M)\frac{\sigma^2}{\tau_{min}} + \varphi^2(\tau_{\min}) \right).$$

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

$$\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau)}}{\tau} + \varphi^2(\tau) \geq \sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} \geq 1$$

$$\geq \frac{1}{2}\left(\sqrt{\log(M)}\sigma^2 \frac{\sqrt{\mathcal{N}(\tau_{\min})}}{\tau_{\min}} + \varphi^2(\tau_{\min}) \right),$$

which proves the assertion. $\qquad\square$

Remark 5.2. For given $\sigma > 0$ the condition from (5.4) can always be satisfied. Below we shall further specify this as follows. If $\tau_{\max}$ is chosen as $\|T^*T\|$ then $\mathcal{N}(\tau_{\max}) \geq 1/2$, such 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 the condition (5.4) 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 5.2. *Suppose that M and $\tau_{\min}$ are chosen such that (5.4) 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 5.1. *Let α, β be fixed and $\Phi_\alpha^\star$ the test defined in (5.3). Suppose that $\tau_{\max} = \|T^*T\|$, $\tau_{\min}$ is chosen such that $M \geq \log_{1/q}\left(\sqrt{2}\|T^*T\| \right) + \log_{1/q}(1/\sigma^2)$. Let τ_* be given from*

$$(5.5) \qquad \varphi^2(\tau_*) = \sigma^2 \sqrt{\log\log_{1/q}(\frac{1}{\sigma^2})} \frac{\sqrt{\mathcal{N}(\tau_*)}}{\tau_*},$$

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

$$(5.6) \qquad \frac{\log\log(\frac{1}{\sigma^2})}{\mathcal{N}(\tau_*)} = o(1) \quad as \quad \sigma \to 0,$$

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

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

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

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

We shall indicate that the assumption (5.6) is valid in many cases.

Lemma 5.4. *If there is a constant $c > 0$ such that the effective dimension obeys*

$$(5.7) \qquad\qquad \mathcal{N}(\tau) \geq c \log(1/\tau),$$

and if the smoothness increases at least as

$$(5.8) \qquad\qquad \varphi(\tau) \leq \left(\log \log_{1/q}(1/\tau) \right)^4,$$

as $\tau \to 0$, then (5.6) is valid.

Proof. The parameter τ_* is determined from (5.5), and under (5.8) we find that

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

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

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

as σ, and hence τ_*, tend to zero. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Remark 5.3. This result covers many of the interesting cases, in particular the ones from Examples 3 & 4. In these cases Theorem 5.1 exhibits that the separation radii obey

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

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

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

Remark 5.4. 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

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

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. Inequalities for Gaussian elements in Hilbert space

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

$$P\left(\|X\|^2 - \mathbb{E}\|X\|^2 \geq x^2 + 2x\sqrt{\mathbb{E}\|X\|^2} \right) \leq \exp\left(-\frac{x^2}{2v^2} \right).$$

where

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

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

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

Hence, we get

$$
\begin{aligned}
P\left(\|X\|^2 - \mathbb{E}\|X\|^2 \geq x^2 + 2x\sqrt{\mathbb{E}\|X\|^2}\right) &\leq P\left(\|X\|^2 \geq \left(\mathbb{E}[\|X\|] + x\right)^2\right), \\
&= P\left(\|X\| \geq \mathbb{E}[\|X\|] + x\right), \\
&\leq \exp\left(-\frac{x^2}{2v^2}\right),
\end{aligned}
$$

where for the last inequality we have used [19, Lemma 3.1]. $\qquad\square$

Lemma A.2. *Let RY be as in (2.1). Then*

$$P_f\left(\|RY\|^2 - \mathbb{E}_f\|RY\|^2 \leq -2\sqrt{\Sigma x_\beta}\right) \leq \beta,$$

where Σ is from (2.10).

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

References

[1] Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

[2] Yannick Baraud, Sylvie Huet, and Béatrice Laurent. Adaptive tests of linear hyptoheses by model selection. *Ann. of Stat.*, 31:225–251, 2003.

[3] Nicolai Bissantz, Thorsten Hohage, Axel Munk, and Frits H. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, 45(6):2610–2636, 2007.

[4] Gilles Blanchard and Peter Mathé. Discrepancy principle for statistical inverse problems with application to conjugate gradient regularization. *Inverse Problems*, 28(11):pp. 115011, 2012.

[5] Ranko Bojanic and Eugene Seneta. A unified theory of regularly varying sequences. *Math. Z.*, 134:91–106, 1973.

[6] Andrea Caponnetto. Optimal rates for regularization operators in learning theory. Technical Report CSAIL-TR 2006-062, Massachusetts Institute of Technology, 2006.

[7] Laurent Cavalier. Inverse problems in statistics. In *Inverse problems and high-dimensional estimation*, volume 203 of *Lect. Notes Stat. Proc.*, pages 3–96. Springer, Heidelberg, 2011.

[8] Carl de Boor. Bounding the error in spline interpolation. *SIAM Rev.*, 16:531–544, 1974.

[9] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[10] Magalie Fromont and Béatrice Laurent. Adaptive goodness-of-fit tests in a density model. *Ann. of Stat.*, 34:1–45, 2006.

[11] Bernd Hofmann and Peter Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J. Numer. Anal.*, 45(3):1122–1141 (electronic), 2007.

[12] Yuri I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.*, 2(2):85–114, 1993.

[13] Yuri I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.*, 2(3):171–189, 1993.

[14] Yuri I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.*, 2(4):249–268, 1993.

[15] Yuri I. Ingster, Teofanis Sapatinas, and Irina A. Suslina. Minimax signal detection in ill-posed inverse problems. *Ann. Statist.*, 40:1524–1549, 2012.

[16] Qinian Jin and Peter Mathé. Oracle inequality for a statistical Raus–Gfrerer type rule. arXiv:1206.3101v1 [math.NA],`http://arxiv.org/pdf/1206.3101`, 2012.

[17] Béatrice Laurent, Jean-Michel Loubes, and Clément Marteau. Testing inverse problems: a direct or an indirect problem? *J. Statist. Plann. Inference*, 141(5):1849–1861, 2011.

[18] Béatrice Laurent, Jean-Michel Loubes, and Clément Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.*, 6:91–122, 2012.

[19] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[20] Peter Mathé and Bernd Hofmann. How general are general source conditions? *Inverse Problems*, 24(1):015009, 5, 2008.

[21] Peter Mathé and Nadine Schöne. Regularization by projection in variable Hilbert scales. *Appl. Anal.*, 87(2):201–219, 2008.

[22] Peter Mathé and Ulrich Tautenhahn. Interpolation in variable Hilbert scales with application to inverse problems. *Inverse Problems*, 22(6):2271–2297, 2006.

[23] Frank Natterer. Regularisierung schlecht gestellter Probleme durch Projektionsverfahren. *Numer. Math.*, 28(3):329–341, 1977.

[24] Vladimir Spokoiny. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24:2477–2498, 1996.

[25] Gennadi M. Vaĭnikko and Uno 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.

Institut Mathématiques de Toulouse, INSA de Toulouse, Université de Toulouse
E-mail address: `clement.marteau@math.univ-toulouse.fr`

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany
E-mail address: `peter.mathe@wias-berlin.de`