A PROBABILISTIC ORACLE INEQUALITY AND QUANTIFICATION OF UNCERTAINTY OF A MODIFIED DISCREPANCY PRINCIPLE FOR STATISTICAL INVERSE PROBLEMS[*]

TIM JAHN[†]

Abstract. In this note we consider spectral cut-off estimators to solve a statistical linear inverse problem under arbitrary white noise. The truncation level is determined with a recently introduced adaptive method based on the classical discrepancy principle. We provide probabilistic oracle inequalities together with quantification of uncertainty for general linear problems. Moreover, we compare the new method to existing ones, namely the early stopping sequential discrepancy principle and the balancing principle, both theoretically and numerically.

1. Introduction. In this note we take a closer look at a recently introduced modified discrepancy principle for solving an inverse problem by means of spectral cut-off. The problem of interest reads

$$(1.1) \qquad\qquad Kx = y,$$

where $K : \mathcal{X} \to \mathcal{Y}$ is a compact injective operator with dense range between infinite-dimensional Hilbert spaces. The problem (1.1) is known to be ill-posed in the sense that K is not continuously invertible on the whole space. This causes problems, since the exact right-hand side $y^\dagger \in \mathcal{R}(K) \subset \mathcal{Y}$ is unknown and we just have access to component measurements corrupted by noise. For the abstract corrupted data, we write

$$y^\delta = y^\dagger + \delta Z,$$

where $\delta > 0$ is the noise level and Z is centred white noise with finite second moments, i.e., it holds that

(i) $\mathbb{E}[(Z, y)] = 0,$
(ii) $\mathbb{E}[(Z, y)(Z, y')] = (y, y'),$
(iii) $(Z, y) \overset{d}{=} \dfrac{\|y\|}{\|y'\|}(Z, y'),$
for all $y, y' \in \mathcal{Y}.$

We have to give an approximation to the true solution $x^\dagger$ based on component measurements $(y^\delta, y_1), (y^\delta, y_2), \dots$ with $y_1, y_2, \dots \in \mathcal{Y}$. Under the above assumptions the forward operator K can be fully described by its singular value decomposition. There exist orthonormal bases $(v_j)_{j \in \mathbb{N}} \subset \mathcal{X}$ and $(u_j)_{j \in \mathbb{N}} \subset \mathcal{Y}$ as well as a sequence $\sigma_1 \geq \sigma_2 \geq \cdots > 0$ converging to zero such that $Kv_j = \sigma_j u_j$ and $K^* u_j = \sigma_j v_j$ for all $j \in \mathbb{N}$. We define the spectral cut-off estimator of $x^\dagger$ via

$$x_k^\delta := \sum_{j=1}^k \frac{(y^\delta, u_j)}{\sigma_j} v_j,$$

with the truncation level k to be determined from the data y^δ and the noise level δ. Note that here we assume the noise level δ to be known in advance; see [14, 15, 30] on how to estimate it in a general setting by a generic method. It is well known that spectral cut-off estimators have excellent theoretical properties, but may be infeasible for general problems in very high dimension due to the fact that the singular value decomposition is unknown in practice and has to be calculated with enormous costs. However, recently, computationally efficient methods like the randomised singular value decomposition [16] have been analysed in the context of regularisation theory and show the potential to allow the use of spectral cut-off estimators in practically relevant settings.

Choosing a suitable truncation level k is one of the main issues in regularisation theory, and plenty of different techniques have been analysed in the past. For general a priori error bounds depending on unknown properties of the exact solution $x^\dagger$, we refer to [4]. Of special interest are adaptive (see [10, 34]) a posteriori methods, which are strategies to choose k dependent on only the noisy measurement y^δ and the noise level δ. Many of the parameter choice rules in statistical inverse problems are adapted from classical statistical methods used for direct regression problems, i.e., problems where K is the identity in (1.1). We name here empirical risk minimisation [21], the balancing principle [29] and generalised cross-validation [24], which are based on Stein's unbiased risk estimation [32], Lepski's method [20] and cross-validation [36], respectively. Others have their roots in the classical deterministic regularisation theory, e.g., the famous discrepancy principle [31] or heuristic methods like the quasi-optimality criterion [18, 33] or the L-curve method [12].

In [17] we proposed a modification of the discrepancy principle for a data-driven choice of k. Originally, the discrepancy principle had its root in the classical deterministic theory, where one assumes that one has an absolute upper bound on the norm of noise $\|y^\delta - y^\dagger\|$. It then follows the paradigm that the data y^δ should only be approximated up to the amount of noise, i.e., k should be determined such that $\|Kx_k^\delta - y^\delta\| \approx \|y^\delta - y^\dagger\|$. In the white noise setting, however, it holds that $\mathbb{E}\|y^\delta - y^\dagger\|^2 = \sum_{j=1}^\infty \delta^2 \mathbb{E}(Z, u_j)^2 = \infty$, and therefore the classical discrepancy principle is not applicable. Consequently the discrepancy principle has to be adapted adequately to the white noise case. This has been done in the past by either pre-smoothing the problem (1.1) (see [7, 22]) or by working directly on a finite-dimensional discrete problem [5, 35]. In the recent work [17], we proposed a modified discrepancy principle using discretisation.

However, different from the above-mentioned works, here the discretisation dimension is treated as an additional regularisation parameter. Precisely, we discretise (1.1) and consider only the first m components $(y^\delta, u_1), \ldots, (y^\delta, u_m)$. Then the data error fulfills

$$\mathbb{E}\left[\left\|\sum_{j=1}^m (y^\delta - y^\dagger, u_j)u_j\right\|^2\right] = m\delta^2,$$

and the classical implementation of the discrepancy principle yields

$$(1.2) \qquad k_{\mathrm{dp}}^\delta(m) := \left\{ k \geq 0 : \sqrt{\sum_{j=k+1}^m (y^\delta, u_j)^2} \leq \tau\sqrt{m}\delta \right\},$$

where $\tau > 1$ is a fudge parameter. It remains to determine the discretisation level m. In [17] it was shown that the ultimate choice

$$(1.3) \qquad k_{\mathrm{dp}}^\delta = \max_{m \in \mathbb{N}} k_{\mathrm{dp}}^\delta(m)$$

yields a convergent regularisation method, as the noise level δ tends to zero.

In this article we give a more complete picture and generalise the main result from [17] to arbitrary K and prove an oracle-type inequality [9] with a controlled probability, instead of asymptotic convergence rates under source conditions [19, 26]. Oracle inequalities have become popular in statistics and guarantee that, after fixing a family of estimators (in our case spectral cut-off estimators, where the estimators are indexed by the truncation level k), one obtains, up to a constant, the optimal error for a given exact solution $x^\dagger$. They depend strongly on the fact that, under white noise, one has sharp estimates for the data propagation error (variance). Minimax convergence rates over source sets, on the other hand, are classic in the deterministic theory, where one only knows an upper bound of the norm of the error, without any structural information. They guarantee optimal convergence in a worst-case fashion where the true solution is an element of an unknown source set. See the survey article [8] for more details. We state here a special case of the general main result presented in the following Section 2.

COROLLARY 1.1. *Assume that there exist $q, c_q, C_q > 0$ such that $C_q j^{-q} \geq \sigma_j^2 \geq c_q j^{-q}$ for all $j \in \mathbb{N}$ and assume that the white noise Z has finite fourth moment $\mathbb{E}(Z, y)^4 \leq \|y\|^4 \gamma_4 < \infty$. Then, for all $\kappa \geq 3$, the following holds:*

$$\sup_{\substack{x^\dagger \in \mathcal{X} \\ \bar{k}^\delta_{\mathrm{pr}}(x^\dagger) \geq \kappa}} \mathbb{P}\Big(\|x^\delta_{k^\delta_{\mathrm{dp}}} - x^\dagger\| \leq C_\tau \min_{k \in \mathbb{N}} \|x^\delta_k - x^\dagger\| \Big)$$

$$\geq 1 - \max\left(\frac{12}{\tau^2 + 2\tau - 3}, 9 \right) \sqrt{\frac{2(1 + \gamma_4)}{\kappa}} - \frac{2(\gamma_4 + 1)\bar{C}_q}{\kappa},$$

with

$$C_\tau := \max\left(\sqrt{2}\Big(\frac{\tau + 1}{\tau - 1} + 1 \Big), \sqrt{2} + (2\tau + 1)\sqrt{\frac{9^{1+q}(1 + q)C_q}{c_q}4} \right)$$

and $\bar{C}_q$ as given in Section 4.

The quantity $\bar{k}^\delta_{\mathrm{pr}}(x^\dagger)$ will be defined in the next section and ensures that $x^\dagger$ can be distinguished sufficiently from the zero signal, relative to the noise level δ. In particular, it holds that $\bar{k}^\delta_{\mathrm{pr}}(x^\dagger) \to \infty$ as $\delta \to 0$ for all non-degenerate $x^\dagger$ (for which arbitrarily large $j \in \mathbb{N}$ exist with $(x^\dagger, u_j) \neq 0$). To understand this condition better, consider the extreme case $x^\dagger = 0$. Here the minimal error could only be obtained for $k^\delta_{\mathrm{dp}} = 0$, which obviously cannot occur with probability tending to 1 (as, e.g., $\delta \to 0$), since a finite fixed number of (random) measurements cannot be controlled with probability tending to 1. Thus the corollary states that, for mildly ill-posed problems, we obtain, up to a constant, the best possible error for spectral cut-off regularisation, uniformly over all signals with sufficiently many relevant components.

In Section 3 we comment on points that were left open in [17], e.g., on how to perform the maximisation over the infinite set $\mathbb{N}$ practically. We end this introduction by stressing that the main difference of the proposed modified discrepancy principle from existing methods is the minimality of the assumptions. In Theorems 2.2 and 2.4 the white noise is only assumed to have a finite second moment and also the singular values of K are arbitrary. Numerical and theoretical comparisons between different regularisation methods for statistical inverse problems can be found in, e.g., [2, 23, 37]. In Sections 3 and 5 we compare our new method to the parameter choice strategies to which it is most closely related, namely, the early stopping discrepancy principle [5, 6] and the balancing principle [1, 25, 27]. In particular, we show that, similar to the balancing principle, the modified discrepancy principle also coincides with Lepski's method in the direct case $K = \mathrm{Id}$ and discuss how the different viewpoint yields an early-stopping-type implementation.

2. Main results. For a truncation level k, we decompose the error into two parts, a data propagation error (variance) and an approximation error (bias):

$$
\begin{aligned}
\|x_k^\delta - x^\dagger\|^2 &= \sum_{j=1}^{k} \left(\frac{(y^\delta, u_j)}{\sigma_j} - (x^\dagger, v_j) \right)^2 + \sum_{j=k+1}^{\infty} (x^\dagger, v_j)^2 \\
&= \sum_{j=1}^{k} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} + \sum_{j=k+1}^{\infty} (x^\dagger, v_j)^2 \\
&= \|x_k^\delta - \mathbb{E}[x_k^\delta]\|^2 + \|\mathbb{E}[x_k^\delta] - x^\dagger\|^2.
\end{aligned}
$$

For the analysis, it will be convenient to consider not only the strong error $\|x_k^\delta - x^\dagger\|$, but also the predictive (or weak) error $\|K(x_k^\delta - x^\dagger)\|$, for which we obtain

$$
\|K(x_k^\delta - x^\dagger)\|^2 = \sum_{j=1}^{k} (y^\delta - y^\dagger, u_j)^2 + \sum_{j=k+1}^{\infty} (y^\dagger, u_j)^2;
$$

note that $(y^\dagger, u_j) = \sigma_j (x^\dagger, v_j)$. Thus we can split the error into two parts, one part that increases monotonically with k and another that decreases monotonically with k.

Now note that minimising such a sum of a monotonically increasing and a monotonically decreasing term is in essence equivalent to balancing the two terms. We thus define the following two truncation levels:

$$
k_{\mathrm{pr}}^\delta(x^\dagger) := \min \left\{ k \in \mathbb{N}_0 : \sum_{j=1}^{k} (y^\delta - y^\dagger, u_j)^2 \geq \sum_{j=k+1}^{\infty} (y^\dagger, u_j)^2 \right\},
$$

$$
k_{\mathrm{st}}^\delta(x^\dagger) := \min \left\{ k \in \mathbb{N}_0 : \sum_{j=1}^{k} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} \geq \sum_{j=k+1}^{\infty} (x^\dagger, v_j)^2 \right\}.
$$

They balance the competing error terms in strong and predictive norm, respectively. For the sake of self-containedness, we formulate the aforementioned well-known result (see, e.g., Lemma 3.2 in [3]), which states that k_{pr}^δ or $k_{\mathrm{pr}}^\delta - 1$ and k_{st}^δ or $k_{\mathrm{st}}^\delta - 1$ (up to a constant factor) minimise the corresponding error norms.

PROPOSITION 2.1. *For all $x^\dagger \in \mathcal{X}$ with $k_{\mathrm{pr}}^\delta(x^\dagger) \geq 1$, the following hold:*

$$
\min_{k \in \mathbb{N}_0} \|K(x_k^\delta - x^\dagger)\| \geq \frac{1}{\sqrt{2}} \min \left(\|K(x_{k_{\mathrm{pr}}^\delta}^\delta - x^\dagger)\|, \|K(x_{k_{\mathrm{pr}}^\delta - 1}^\delta - x^\dagger)\| \right),
$$

$$
\min_{k \in \mathbb{N}_0} [\|x_k^\delta - x^\dagger\|] \geq \frac{1}{\sqrt{2}} \min \left(\|x_{k_{\mathrm{st}}^\delta}^\delta - x^\dagger\|, \|x_{k_{\mathrm{st}}^\delta - 1}^\delta - x^\dagger\| \right).
$$

Proof. We use the convention $\sum_{j=1}^{0} = 0$. By monotonicity and the definition of k_{pr}^δ, the

following hold for $k \geq k_{\mathrm{pr}}^{\delta}$,

$$\|K(x_k^{\delta} - x^{\dagger})\|^2 = \sum_{j=1}^{k}(y^{\delta} - y^{\dagger}, u_j)^2 + \sum_{j=k+1}^{\infty}(y^{\dagger}, u_j)^2 \geq \sum_{j=1}^{k}(y^{\delta} - y^{\dagger}, u_j)^2$$

$$\geq \frac{1}{2}\left(\sum_{j=1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta} - y^{\dagger}, u_j)^2 + \sum_{j=1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta} - y^{\dagger}, u_j)^2\right)$$

$$\geq \frac{1}{2}\left(\sum_{j=1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta} - y^{\dagger}, u_j)^2 + \sum_{j=k_{\mathrm{pr}}^{\delta}+1}^{\infty}(y^{\dagger}, u_j)^2\right) = \frac{1}{2}\|K(x_{k_{\mathrm{pr}}^{\delta}}^{\delta} - x^{\dagger})\|^2,$$

and for $k \leq k_{\mathrm{pr}}^{\delta} - 1$,

$$\|K(x_k^{\delta} - x^{\dagger})\|^2 = \sum_{j=1}^{k}(y^{\delta} - y^{\dagger}, u_j)^2 + \sum_{j=k+1}^{\infty}(y^{\dagger}, u_j)^2 \geq \sum_{j=k+1}^{\infty}(y^{\dagger}, u_j)^2$$

$$\geq \frac{1}{2}\left(\sum_{j=k_{\mathrm{pr}}^{\delta}}^{\infty}(y^{\dagger}, u_j)^2 + \sum_{j=k_{\mathrm{pr}}^{\delta}}^{\infty}(y^{\dagger}, u_j)^2\right)$$

$$\geq \frac{1}{2}\left(\sum_{j=1}^{k_{\mathrm{pr}}^{\delta}-1}(y^{\delta} - y^{\dagger}, u_j)^2 + \sum_{j=k_{\mathrm{pr}}^{\delta}}^{\infty}(y^{\dagger}, u_j)^2\right) = \frac{1}{2}\|K(x_{k_{\mathrm{pr}}^{\delta}-1}^{\delta} - x^{\dagger})\|^2,$$

which proves the assertion for k_{pr}^{δ}. The argumentation for the second assertion is analogous.
$\square$

Note that y^{δ} and hence also k_{pr}^{δ} and k_{st}^{δ} are random quantities. In order to formulate our main results, it is handy to have a deterministic quantity. We thus define

$$\bar{k}_{\mathrm{pr}}^{\delta}(x^{\dagger}) := \min\left\{k \in \mathbb{N}_0 : \delta^2 k \geq \sum_{j=k+1}^{\infty}(y^{\dagger}, u_j)^2\right\}.$$

Obviously $\bar{k}_{\mathrm{pr}}^{\delta}$ is closely related to k_{pr}^{δ}, since $k\delta^2 = \mathbb{E}\left[\sum_{j=1}^{k}(y^{\delta} - y^{\dagger}, u_j)^2\right]$ and thus $\bar{k}_{\mathrm{pr}}^{\delta}$ balances the expected squared prediction error norm $\mathbb{E}\|K(x_k^{\delta} - x^{\dagger})\|^2$. We come to the first main result, a full oracle inequality for the prediction error.

THEOREM 2.2. *For all $\kappa \geq 3$, it holds that*

$$\sup_{\substack{x^{\dagger} \in \mathcal{X} \\ \bar{k}_{\mathrm{pr}}^{\delta}(x^{\dagger}) \geq \kappa}} \mathbb{P}\left(\|K(x_{k_{\mathrm{dp}}^{\delta}}^{\delta} - x^{\dagger})\| \leq C_{\tau} \min_{k \in \mathbb{N}_0}\|K(x_k^{\delta} - x^{\dagger})\|\right)$$

$$\geq 1 - \max\left(\frac{12}{\tau^2 + 2\tau - 3}, 9\right)\mathbb{E}\left[\left|\frac{1}{\kappa}\sum_{j=1}^{\lceil \kappa/3 \rceil}((Z, u_j)^2 - 1)\right|\right] \to 1,$$

as $\kappa \to \infty$, with $C_{\tau} := \sqrt{6}\sqrt{\frac{3}{2}(A_{\tau} + 1) + B_{\tau}}$, and with A_{τ} and B_{τ} as given below in Section 4.

We want to point out the generality of the above result, which perfectly reflects the paradigm that the discrepancy principle balances approximation and data propagation error in the image space, in that it results, up to a constant, in the error of the weak balanced oracle.

However, usually one is interested in the error in the strong norm. The result we obtain here will depend on the difference between the weak balanced oracle k_{pr}^δ and the strong balanced oracle k_{st}^δ. We give an important relation between the two oracles in the next proposition.

PROPOSITION 2.3. *It holds that* $k_{\mathrm{pr}}^\delta \leq k_{\mathrm{st}}^\delta$.

Proof. We have

$$\sum_{j=1}^{k_{\mathrm{pr}}^\delta-1} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} \leq \frac{\sum_{j=1}^{k_{\mathrm{pr}}^\delta-1}(y^\delta - y^\dagger, u_j)^2}{\sigma_{k_{\mathrm{pr}}^\delta-1}^2} < \frac{1}{\sigma_{k_{\mathrm{pr}}^\delta-1}^2} \sum_{j=k_{\mathrm{pr}}^\delta}^\infty (y^\dagger, u_j)^2$$

$$= \frac{1}{\sigma_{k_{\mathrm{pr}}^\delta-1}^2} \sum_{j=k_{\mathrm{pr}}^\delta}^\infty \sigma_j^2 (x^\dagger, v_j)^2 \leq \sum_{j=k_{\mathrm{pr}}^\delta}^\infty (x^\dagger, v_j)^2,$$

where we used the definition of k_{pr}^δ in the second step. By the definition of k_{st}^δ, it follows that $k_{\mathrm{st}}^\delta > k_{\mathrm{pr}}^\delta - 1$, hence $k_{\mathrm{st}}^\delta \geq k_{\mathrm{pr}}^\delta$. $\quad\square$

Proposition 2.3 already indicates that the strong error of the discrepancy principle might be suboptimal in the case that k_{pr}^δ is substantially smaller than k_{st}^δ. This is indeed the case. The next theorem shows that the error in the strong norm depends on the difference between k_{pr}^δ and k_{st}^δ.

THEOREM 2.4. *For $\kappa \geq 3$ it holds that*

$$\sup_{\substack{x^\dagger \in \mathcal{X} \\ \bar{k}_{\mathrm{pr}}^\delta(x^\dagger) \geq \kappa}} \mathbb{P}\left(\|x_{k_{\mathrm{dp}}^\delta}^\delta - x^\dagger\| \leq C_\tau \left(\min_{k \in \mathbb{N}} \|x_k^\delta - x^\dagger\| + \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{k_{\mathrm{st}}^\delta} (x^\dagger, v_j)^2} \right) \right)$$

$$\geq 1 - \max\left(\frac{12}{\tau^2 + 2\tau - 3}, 9 \right) \mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^{\lceil \kappa/3 \rceil} ((Z, u_j)^2 - 1) \right| \right] \to 1$$

as $\kappa \to \infty$, with

$$C_\tau := \sqrt{2} \max\left(\frac{\tau+1}{\tau-1} + 1, 1 + \sqrt{\frac{3}{8}}(3\tau + 1) \right).$$

Under additional assumptions, we can bound the probability in Theorems 2.2 and 2.4 explicitly.

PROPOSITION 2.5. *Assume that, for some $2 < p \leq 4$, the white noise has finite pth moment, i.e., $\mathbb{E}|(Z, y)|^p = \gamma_p \|y\| < \infty$. Then*

$$\mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^\kappa ((Z, u_j)^2 - 1) \right| \right] \leq 2^{1-2/p} (\gamma_p + 1)^{2/p} \kappa^{2/p-1}.$$

Proof. By convexity and the Marcinkiewicz–Zygmund inequality (Corollary 8.2 in [11]), it holds that

$$\mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^\kappa ((Z, u_j)^2 - 1) \right| \right] \leq \left(\mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^\kappa ((Z, u_j)^2 - 1) \right|^{p/2} \right] \right)^{2/p}$$

$$\leq (\kappa^{1-p/2} \mathbb{E}|(Z, u_1)^2 - 1|^{p/2})^{2/p}$$

$$\leq (2^{p/2-1}(\gamma_p + 1)1)^{2/p} \kappa^{2/p-1}$$

$$\leq 2^{1-2/p}(\gamma_p + 1)^{2/p} \kappa^{2/p-1}. \quad\square$$

Unlike Theorem 2.2, Theorem 2.4 above is not a proper oracle inequality, since the additional term is likely to dominate. In fact, this perfectly illustrates in full generality one well-known drawback of the discrepancy principle, namely that it saturates. Note that this saturation is hidden when one considers convergence rates over source conditions for the discrepancy principle with spectral cut-off. For some operators we can show that the additional term will not dominate the total error (see, e.g., Corollary 1.1), and we obtain a full oracle inequality for the strong error norm. The proofs of Theorem 2.2, Theorem 2.4, and Corollary 1.1 are deferred to Section 4.

3. Comparison to other parameter choice strategies. In this section we will compare our rule to two other parameter choice strategies, namely the balancing principle and the early-stopping discrepancy principle. We start with the balancing principle. This principle is based on ideas introduced by Lepski [20]. Lepski's method is originally formulated in a regression setting. A signal corrupted by white noise has to be recovered. This corresponds to our problem (1.1) with $K = \mathrm{Id}$. We state the problem explicitly. Given noisy component measurements

$$(y^\delta, u_j) = (y^\dagger, u_j) + \delta(Z, u_j), \quad j \in \mathbb{N},$$

and the spectral cut-off estimators

$$y_k^\delta := \sum_{j=1}^k (y^\delta, u_j) u_j,$$

find the truncation level $k \in \mathbb{N}$ such that the error $\|y_k^\delta - y^\dagger\|$ is optimal in some sense. Lepski proposed the following rule:

$$k_{\mathrm{Lep}}^\delta := \min\{k \in \mathbb{N} : \|y_m^\delta - y_k^\delta\| \le \kappa \sqrt{m} \delta, \ \forall \, m > k\},$$

with a fudge parameter $\kappa > 1$. The idea was that k should be large enough such that the differences would be dominated by the variance; note that

$$\mathbb{E}\|y_m^\delta - \mathbb{E}[y_m^\delta]\|^2 = \mathbb{E}\left[\sum_{j=1}^m (y^\delta - y^\dagger, u_j)^2\right] = m\delta^2.$$

In the above direct setting, our approach (1.3) and (1.2) would yield

$$k_{\mathrm{dp}}^\delta := \max_{m \in \mathbb{N}} k_{\mathrm{dp}}^\delta(m) = \max_{m \in \mathbb{N}} \min\left\{0 \le k \le m : \sqrt{\sum_{j=k+1}^m (y^\delta, u_j)^2} \le \tau\sqrt{m}\delta\right\}.$$

It is not hard to show that the two parameter choices actually coincide in this setting.

THEOREM 3.1. *Let* $\kappa = \tau$. *Then* $k_{\mathrm{Lep}}^\delta = k_{\mathrm{dp}}^\delta$ *pointwise.*

Proof. We observe that

$$\sum_{j=k+1}^m (y^\delta, u_j)^2 = \|y_m^\delta - y_k^\delta\|^2$$

and hence

$$k_{\mathrm{dp}}^\delta = \max_{m \in \mathbb{N}}\{0 \le k \le m : \|y_m^\delta - y_k^\delta\| \le \tau\sqrt{m}\delta\}.$$

Now let $k = k_{\mathrm{Lep}}^\delta$. By definition, for all $m > k$ it holds that $\|y_m^\delta - y_k^\delta\| \leq \tau\sqrt{m}\delta$. We deduce $k_{\mathrm{dp}}^\delta(m) \leq k$ for all $m \in \mathbb{N}$ and thus $k_{\mathrm{dp}}^\delta \leq k_{\mathrm{Lep}}^\delta$. Now let $k = k_{\mathrm{Lep}}^\delta - 1$. By definition, there exists $m > k$ with $\|y_m^\delta - y_k^\delta\| > \tau\sqrt{m}\delta$. Thus $k_{\mathrm{dp}}^\delta(m) > k$ and therefore $k_{\mathrm{dp}}^\delta \geq k+1 = k_{\mathrm{Lep}}^\delta$. The proof is finished. $\square$

So we see that the two methods coincide for direct regression problems, i.e., when $K = \mathrm{Id}$. When a forward operator K is involved, the balancing principle for inverse problems deduced from Lepski's method, however, follows a different approach. There, the estimators are compared, not in the weak norm, but in the strong norm. The variance of the spectral cut-off estimator in the strong norm is

$$\mathbb{E}\|x_m^\delta - \mathbb{E}[x_m^\delta]\|^2 = \sum_{j=1}^m \frac{\mathbb{E}(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} = \delta^2 \sum_{j=1}^m \frac{1}{\sigma_j^2},$$

i.e., we have

$$(3.1) \qquad k_{\mathrm{bal}}^\delta := \min\left\{ k \geq 0 : \|x_m^\delta - x_k^\delta\| \leq \kappa\delta\sqrt{\sum_{j=1}^m \frac{1}{\sigma_j^2}}, \ \forall \, m > k \right\}.$$

Here the choice of κ is delicate. The reason for this is the following. In the direct setting we have

$$\|y_m^\delta - \mathbb{E}[y_m^\delta]\|^2 - \mathbb{E}\|y^\delta - \mathbb{E}[y_m^\delta]\|^2 = \sum_{j=1}^m ((y^\delta - y^\dagger, u_j)^2 - 1) = \delta^2 \sum_{j=1}^m ((Z, u_j)^2 - 1),$$

which is an i.i.d. sum of unbiased random variables and hence a reverse martingale; see the proof of Proposition 4.1 in Section 4 below. Therefore, its whole trajectory, i.e., $(\|y_m^\delta - \mathbb{E}[y_m^\delta]\|^2 - \mathbb{E}\|y^\delta - \mathbb{E}[y_m^\delta]\|^2)_{m\in\mathbb{N}}$, can be controlled. Unlike this, in the case of inverse regression, we have

$$\|x_m^\delta - \mathbb{E}[x_m^\delta]\|^2 - \mathbb{E}\|x_m^\delta - \mathbb{E}[x_m^\delta]\|^2 = \delta^2 \sum_{j=1}^m \frac{(Z, u_j)^2 - 1}{\sigma_j^2},$$

which is just a sum with independent summands of increasing variance. In particular, depending on the behaviour of the singular values σ_j, the single last summand may give a significant contribution to the whole sum or even dominate the whole sum. Consequently, one considers only finitely many estimators in (3.1), i.e., $m \leq D$, where D depends on the noise level δ and the spectrum of K, and κ will also depend on D and therefore on δ. Typically one sets $\kappa \sim \log(\delta)^{-1}$ and then obtains optimal convergence in L^2 up to a logarithmic correction for polynomially and exponentially ill-posed problems under Gaussian noise. While the concrete dependence of κ on δ is usually tailored to obtain optimal convergence in L^2 for Gaussian noise, it is not directly clear whether a constant κ could yield convergence in probability; cf. Theorems 2.2 and 2.4. The following counter-example shows that, at least for exponentially ill-posed problems, κ has to depend on δ.

EXAMPLE 3.2. Let $\sigma_j^2 = e^{-j}$ and $x^\dagger = 0$ and let Z be Gaussian white noise. Assume that some $\kappa > 1$ is fixed and set $m_\delta = \lceil \log(\delta^{-2}) \rceil$. Let k_{bal}^δ be determined by the balancing principle (3.1). Then there exists $p_\kappa > 0$ such that

$$(3.2) \qquad \mathbb{P}(\|x_{k_{\mathrm{bal}}^\delta}^\delta - x^\dagger\| \geq 1) \geq p_\kappa$$

for all $0 < \delta \leq e^{-1}$. We show that (3.2) is fulfilled with

$$p_\kappa := \mathbb{P}(|X| > e\kappa),$$

where X is a standard Gaussian. Indeed, we set

$$\Omega_\kappa := \{|(Z, u_{m_\delta})| > e\kappa\}.$$

By definition of p_κ, it holds that $\mathbb{P}(\chi_{\Omega_\kappa}) = p_\kappa$. It also holds that

$$\|x^\delta_{m_\delta} - x^\delta_{m_\delta-1}\|\chi_{\Omega_\kappa} = \delta|(Z, u_{m_\delta})|e^{m_\delta/2}\chi_{\Omega_\kappa} > e\kappa\delta e^{m_\delta/2}$$
$$\geq \kappa\delta\sqrt{\frac{e^{m_\delta+1} - 1}{e - 1}} > \kappa\delta\sqrt{\sum\nolimits_{j=1}^{m_\delta} e^j}$$

since $(e^{m+1} - 1)/(e - 1) \leq ee^m$. Consequently we have

$$k^\delta_{\mathrm{bal}}\chi_{\Omega_\kappa} \geq m_\delta\chi_{\Omega_\kappa}.$$

Ultimately, we find

$$\|x^\delta_{k^\delta_{\mathrm{bal}}} - x^\dagger\|\chi_{\Omega_\kappa} \geq \frac{|(y^\delta, u_{m_\delta})|}{\sigma_{m_\delta}}\chi_{\Omega_\kappa} = \delta e^{m_\delta/2}|(Z, u_{m_\delta})|\chi_{\Omega_\kappa} > e\kappa\chi_{\Omega_\kappa} \geq 1\chi_{\Omega_\kappa}$$

by definition of m_δ, which finishes the proof of the assertion (3.2).

To put it in a nutshell, we saw that both methods are closely related to Lepski's method. The balancing principle has the advantage that the estimators are compared in the same norm in which one wants to have convergence, i.e., the strong norm. In the case of the modified discrepancy principle, the estimators are compared in the weak norm. However, we have much worse control of the variance in the strong norm than in the weak norm, which makes the choice of the fudge parameter more delicate for the balancing principle and slightly deteriorates the rate. Finally, at the end of this section we discuss that the different viewpoint on Lepski's method provided by the modified discrepancy principle allows for another potential benefit related to early stopping.

Note that a drawback of both methods is their high computational costs. A whole series of estimators need to be computed and then compared to each other, before the final choice can be made. A computational very attractive and still convergent method is the early stopping or sequential discrepancy principle [5], which makes use of the computational simplicity of the plain discrepancy principle [28]. Here, a maximal dimension D is chosen first, respectively given by the measurement process. Then the classical discrepancy principle is applied, but with the parameter τ set to one:

$$(3.3) \qquad k^\delta_{\mathrm{es}} := \min\left\{0 \leq k \leq D : \sqrt{\sum\nolimits_{j=k+1}^{D}(y^\delta, u_j)^2} \leq \sqrt{D}\delta\right\}.$$

The name "sequential" or "early stopping" refers to the fact that it takes the estimator that was computed last. Note that usually the singular value decomposition is unknown and has to be approximated numerically. In particular, calculations of later singular vectors are more costly and less accurate than calculations of the first ones. Thus a big advantage is due to the fact that, in order to compute k^δ_{es}, one just needs the first k^δ_{es} singular vectors and values. This can be seen as follows.

In an application we usually have a discretised $K \in \mathbb{R}^{D \times D'}$ with D and D' large, and right-hand side $y^\delta_D \in \mathbb{R}^D$. Expressing y^δ_D in the singular basis of K then gives $y^\delta_D = \sum_{j=1}^{D}(y^\delta, u_j)u_j$ and thus $\sum_{j=k+1}^{D}(y^\delta, u_j)^2 = \|y^\delta_D\|^2 - \sum_{j=1}^{k}(y^\delta, u_j)^2$. Note hereby that the total norm $\|y^\delta_D\|^2$ can be computed without knowledge of singular vectors. Similar to k^δ_{dp}, also k^δ_{es} mimics the predictive oracle k^δ_{pr}. In [5] oracle inequalities for k^δ_{es} are proven.

Owing to the large variance in the residuals, they do not hold for very smooth solutions. Also higher orders of the error distribution are needed; see, e.g., Theorem 4.2 in [15]. We will see in Section 5 that k_{es}^δ tends to be numerically less stable than the other two methods.

Finally, there is an obvious way to combine the early stopping discrepancy principle with the modified discrepancy principle in that one can use the early discrepancy principle to determine a maximum m over which we maximise in (1.2). That is, one could consider

$$k_{\mathrm{com}}^\delta := \max_{m \leq k_{\mathrm{es}}^\delta} k_{\mathrm{dp}}^\delta(m)$$

or, more generally, and to avoid the instabilities of k_{es}^δ, one could set, for $\tau > \tau_{\min} \geq 1$,

$$k_{\mathrm{com}}^\delta := \max_{m \leq m_{\max}} \min \left\{ 0 \leq k \leq m : \sqrt{\sum\nolimits_{j=k+1}^m (y^\delta, u_j)^2} \leq \tau \sqrt{m} \delta \right\},$$

with

$$m_{\max} = \min \left\{ 0 \leq k \leq D : \sqrt{\sum\nolimits_{j=k+1}^D (y^\delta, u_j)^2} \leq \tau_{\min} \sqrt{D} \delta \right\}.$$

4. Proofs. A central tool for the proofs will be the following proposition already used in [17], which allows one to control the measurement error.

PROPOSITION 4.1. *For any $\epsilon > 0$ and $\kappa \in \mathbb{N}$, the following holds:*

$$\mathbb{P}\left(\left| \sum_{j=1}^m (y^\delta - y^\dagger, u_j)^2 - m\delta^2 \right| \geq \varepsilon m \delta^2, \ \forall \, m \geq \kappa \right)$$

$$\leq \frac{1}{\varepsilon} \mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^\kappa ((Z, u_j)^2 - 1) \right| \right] \to 0$$

as $\kappa \to \infty$.

Proof. We give a short proof in order to keep the paper self-contained; see Proposition 3.1 in [17] for more details. It holds that

$$\mathbb{P}\left(\left| \sum_{j=1}^m (y^\delta - y^\dagger, u_j)^2 - m\delta^2 \right| \leq \varepsilon m \delta^2, \ \forall \, m \geq \kappa \right) = \mathbb{P}\left(\sup_{m \geq \kappa} \left| \frac{1}{m} \sum_{j=1}^m X_j \right| \leq \varepsilon \right),$$

where

$$X_j := \frac{(y^\delta - y^\dagger, u_j)^2}{\delta^2} - 1 = (Z, u_j)^2 - 1.$$

Note that the right-hand side of the above equation is independent of the noise level δ. Since $(X_j)_{j \in \mathbb{N}}$ is i.i.d. with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}|X_1| \leq 2$, the sample mean $\left((1/m) \sum_{j=1}^m X_j \right)_{m \in \mathbb{N}}$ is a reverse martingale. Thus the Kolmogorov–Doob inequality (Theorem 16.2 in [11]) yields

$$\mathbb{P}\left(\sup_{m \geq \kappa} \left| \frac{1}{m} \sum_{j=1}^m X_j \right| > \varepsilon \right) \leq \frac{1}{\varepsilon} \mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^\kappa X_j \right| \right] \to 0$$

as $\kappa \to \infty$, by the law of large numbers and the fact that reverse martingales are uniformly integrable. □

Now we can prove the main results. So let $x^\dagger$ be given with $\bar{k}_{\mathrm{pr}}^\delta(x^\dagger) \geq \kappa$. We suppress the dependence on $x^\dagger$ of the balancing index $\bar{k}_{\mathrm{pr}}^\delta(x^\dagger)$ in the following. We carry out the analysis

on the following sequence of events on which we have perfect control of the measurement error:

$$(4.1) \quad \Omega_\kappa := \left\{ \left| \sum_{j=1}^m (y^\delta - y^\dagger, u_j)^2 - m\delta^2 \right| \le \min\left(\frac{(\tau+1)^2}{4} - 1, \frac{1}{3} \right) m\delta^2, \ \forall \, m \ge \frac{\kappa}{3} \right\}.$$

By Proposition 4.1 it holds that

$$(4.2) \qquad \mathbb{P}(\Omega_\kappa) \ge 1 - \max\left(\frac{12}{\tau^2 + 2\tau - 3}, 9 \right) \mathbb{E}\left[\left| \frac{1}{\kappa} \sum_{j=1}^{\lceil \kappa/3 \rceil} ((Z, u_j)^2 - 1) \right| \right].$$

Obviously on Ω_κ we have

$$\frac{2}{3} m\delta^2 \chi_{\Omega_\kappa} \le \sum_{j=1}^m (y^\delta - y^\dagger, u_j)^2 \chi_{\Omega_\kappa} \le \frac{4}{3} m\delta^2 \chi_{\Omega_\kappa},$$

$$\sqrt{\sum_{j=1}^m (y^\delta - y^\dagger, u_j)^2} \, \chi_{\Omega_\kappa} \le \frac{\tau+1}{2} \sqrt{m}\delta \chi_{\Omega_\kappa}.$$

We first show that

$$(4.3) \qquad (k_{\mathrm{pr}}^\delta - 1)\chi_{\Omega_\kappa} \ge \frac{\kappa}{3} \chi_{\Omega_\kappa}$$

for all $\kappa \ge 3$. Indeed, since $\lceil \bar{k}_{\mathrm{pr}}^\delta/3 \rceil \ge \kappa/3$ by definition of Ω_κ, we have that

$$\sum_{j=1}^{\lceil \bar{k}_{\mathrm{pr}}^\delta/3 \rceil} (y^\delta - y^\dagger, u_j)^2 \chi_{\Omega_\kappa} \le \frac{4}{3} \left\lceil \frac{\bar{k}_{\mathrm{pr}}^\delta}{3} \right\rceil \delta^2 \le (\bar{k}_{\mathrm{pr}}^\delta - 1)\delta^2 < \sum_{j=\bar{k}_{\mathrm{pr}}^\delta}^\infty (y^\dagger, u_j)^2$$

$$< \sum_{j=\lceil \bar{k}_{\mathrm{pr}}^\delta/3 \rceil + 1}^\infty (y^\dagger, u_j)^2,$$

and thus $k_{\mathrm{pr}}^\delta \chi_{\Omega_\kappa} > \lceil \bar{k}_{\mathrm{pr}}^\delta/3 \rceil \chi_{\Omega_\kappa}$ by definition, which proves the claim (4.3).

Proof of Theorem 2.2. We begin with an upper bound for k_{dp}^δ. For $A_\tau = ((\tau+1)/(\tau-1))^2$ one has

$$(4.4) \qquad k_{\mathrm{dp}}^\delta \chi_{\Omega_\kappa} \le A_\tau k_{\mathrm{pr}}^\delta.$$

To prove (4.4) we have to show that $k_{\mathrm{dp}}^\delta(m) \le A_\tau k_{\mathrm{pr}}^\delta$ for all $m \in \mathbb{N}$. This is clearly true for all $m \in \mathbb{N}$ with $k_{\mathrm{dp}}^\delta(m) \le k_{\mathrm{pr}}^\delta$. Now consider $m \in \mathbb{N}$ with $k_{\mathrm{dp}}^\delta(m) > k_{\mathrm{pr}}^\delta$. By the defining relation of the discrepancy principle and the definition of k_{pr}^δ and $\chi_{\tilde{\Omega}_{\bar{k}_{\mathrm{st}}^\delta}}$ it holds that

$$\tau \sqrt{m}\delta \chi_{\Omega_\kappa} \le \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta(m)}^m (y^\delta, u_j)^2} \, \chi_{\Omega_\kappa}$$

$$\le \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta(m)}^m (y^\delta - y^\dagger, u_j)^2} \, \chi_{\Omega_\kappa} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta(m)}^m (y^\dagger, u_j)^2} \, \chi_{\Omega_\kappa}$$

$$\le \frac{\tau+1}{2} \sqrt{m}\delta + \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta+1}^m (y^\dagger, u_j)^2}$$

$$\leq \frac{\tau+1}{2}\sqrt{m}\delta + \sqrt{\sum\nolimits_{j=1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta}-y^{\dagger},u_j)^2}\,\chi_{\Omega_{\kappa}}$$

$$\leq \frac{\tau+1}{2}\sqrt{m}\delta + \frac{\tau+1}{2}\sqrt{k_{\mathrm{pr}}^{\delta}}\delta.$$

We obtain

$$\sqrt{m}\delta\chi_{\Omega_{\kappa}} \leq \frac{\tau+1}{\tau-1}\sqrt{k_{\mathrm{pr}}^{\delta}}\delta = \sqrt{A_{\tau}k_{\mathrm{pr}}^{\delta}}\delta.$$

This together with the obvious inequality $k_{\mathrm{dp}}^{\delta}(m) \leq m$ finishes the proof of the claim (4.4).

This shows the stability of the method in the sense that we do not stop too late. To finish the proof, we have to show that we do not stop too early. We claim that

$$(4.5) \qquad \sum_{j=k_{\mathrm{dp}}^{\delta}+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}} \leq B_{\tau}k_{\mathrm{pr}}^{\delta}\delta^2$$

with $B_{\tau} = (3\tau+1)^2/4$. To prove (4.5) let us consider $m = k_{\mathrm{pr}}^{\delta}$. Then, by definition of the discrepancy principle,

$$\tau\sqrt{k_{\mathrm{pr}}^{\delta}}\delta\chi_{\Omega_{\kappa}} \geq \sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^{\delta}(k_{\mathrm{pr}}^{\delta})+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta},u_j)^2}\,\chi_{\Omega_{\kappa}}$$

$$\geq \sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^{\delta}(k_{\mathrm{pr}}^{\delta})+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\dagger},u_j)^2}\,\chi_{\Omega_{\kappa}} - \sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^{\delta}(k_{\mathrm{pr}}^{\delta})+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta}-y^{\dagger},u_j)^2}\,\chi_{\Omega_{\kappa}}$$

$$\geq \sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^{\delta}(k_{\mathrm{pr}}^{\delta})+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\dagger},u_j)^2}\,\chi_{\Omega_{\kappa}} - \frac{\tau+1}{2}\sqrt{k_{\mathrm{pr}}^{\delta}}\delta.$$

By monotonicity and with the convention $\sum_{j=m}^{n} = 0$ for $m > n$, we obtain

$$\sum_{j=k_{\mathrm{dp}}^{\delta}+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}} \leq \sum_{j=k_{\mathrm{dp}}^{\delta}(k_{\mathrm{pr}}^{\delta})+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\dagger},u_j)\chi_{\Omega_{\kappa}} \leq \frac{(3\tau+1)^2}{4}k_{\mathrm{pr}}^{\delta}\delta^2 = B_{\tau}k_{\mathrm{pr}}^{\delta}\delta^2$$

and (4.5) is proven. We combine the two estimates (4.4) and (4.5) to give

$$\|K(x_{k_{\mathrm{dp}}^{\delta}}^{\delta}-x^{\dagger})\|^2\chi_{\Omega_{\kappa}}$$

$$= \sum_{j=1}^{k_{\mathrm{dp}}^{\delta}}(y^{\delta}-y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}} + \sum_{j=k_{\mathrm{dp}}^{\delta}+1}^{\infty}(y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}}$$

$$= \sum_{j=1}^{k_{\mathrm{dp}}^{\delta}}(y^{\delta}-y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}} + \sum_{j=k_{\mathrm{dp}}^{\delta}+1}^{k_{\mathrm{pr}}^{\delta}}(y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}} + \sum_{j=k_{\mathrm{pr}}^{\delta}+1}^{\infty}(y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}}$$

$$\leq \sum_{j=1}^{A_{\tau}k_{\mathrm{pr}}^{\delta}}(y^{\delta}-y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}} + B_{\tau}k_{\mathrm{pr}}^{\delta}\delta^2\chi_{\Omega_{\kappa}} + \sum_{j=1}^{k_{\mathrm{pr}}^{\delta}}(y^{\delta}-y^{\dagger},u_j)^2\chi_{\Omega_{\kappa}}$$

$$(4.6) \qquad \leq \left(\frac{3}{2}(A_{\tau}+1)+B_{\tau}\right)k_{\mathrm{pr}}^{\delta}\delta^2\chi_{\Omega_{\kappa}}.$$

On the other hand we have

$$\|K(x_{k_{\mathrm{pr}}^{\delta}-1}^{\delta} - x^{\dagger})\|^2 \chi_{\Omega_\kappa} = \sum_{j=1}^{k_{\mathrm{pr}}^{\delta}-1} (y^\delta - y^\dagger, u_j)^2 \chi_{\Omega_\kappa} + \sum_{j=k_{\mathrm{pr}}^{\delta}}^{\infty} (y^\dagger, u_j)^2 \chi_{\Omega_\kappa}$$

$$\geq \frac{2}{3}(k_{\mathrm{pr}}^{\delta} - 1)\delta^2 \chi_{\Omega_\kappa} \geq \frac{1}{3} k_{\mathrm{pr}}^{\delta} \chi_{\Omega_\kappa},$$

and also

$$\|K(x_{k_{\mathrm{pr}}^{\delta}}^{\delta} - x^{\dagger})\|^2 = \sum_{j=1}^{k_{\mathrm{pr}}^{\delta}} (y^\delta - y^\dagger, u_j)^2 \chi_{\Omega_\kappa} + \sum_{j=k_{\mathrm{pr}}^{\delta}+1}^{\infty} (y^\dagger, u_j)^2 \chi_{\Omega_\kappa}$$

$$\geq \frac{2}{3} k_{\mathrm{pr}}^{\delta} \delta^2 \chi_{\Omega_\kappa}.$$

Therefore Proposition 2.1 gives

$$\min_{k \in \mathbb{N}} \|K(x_k^\delta - x^\dagger)\| \chi_{\Omega_\kappa} \geq \frac{1}{\sqrt{2}} \min\left(\|K(x_{k_{\mathrm{pr}}^{\delta}}^{\delta} - x^\dagger)\|, \|K(x_{k_{\mathrm{pr}}^{\delta}-1}^{\delta} - x^\dagger)\|\right) \chi_{\Omega_\kappa}$$

$$\geq \frac{1}{\sqrt{6}} \sqrt{k_{\mathrm{pr}}^{\delta}} \delta \chi_{\Omega_\kappa}.$$

Together with (4.6) we obtain

$$\|K(x_{k_{\mathrm{dp}}^{\delta}}^{\delta} - x^\dagger)\| \chi_{\Omega_\kappa} \leq \sqrt{\tfrac{3}{2}(A_\tau + 1) + B_\tau} \sqrt{k_{\mathrm{pr}}^{\delta}} \delta \chi_{\Omega_\kappa} \leq C_\tau \min_{k \geq \mathbb{N}_0} \|K(x_k^\delta - x^\dagger)\|,$$

where

$$C_\tau := \sqrt{6} \sqrt{\tfrac{3}{2}(A_\tau + 1) + B_\tau}.$$

Theorem 2.2 then follows with (4.2). □

Proof of Theorem 2.4. We again resort to the event Ω_κ from (4.1). Note that $k_{\mathrm{st}}^{\delta} \chi_{\Omega_\kappa} \geq k_{\mathrm{pr}}^{\delta} \chi_{\Omega_\kappa} \geq 3\kappa$ by Proposition 2.3. We start by showing the stability of the approach. We claim that

(4.7) $$\sqrt{\sum_{j=1}^{k_{\mathrm{dp}}^{\delta}} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2}} \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^{\delta} \geq k_{\mathrm{st}}^{\delta}\}} \leq \frac{\tau+1}{\tau-1} \sum_{j=k_{\mathrm{dp}}^{\delta}}^{\infty} (x^\dagger, v_j)^2 \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^{\delta} \geq k_{\mathrm{st}}^{\delta}\}}.$$

So assume that $m \geq k_{\mathrm{st}}^{\delta}$ with $k_{\mathrm{dp}}^{\delta}(m) \geq k_{\mathrm{st}}^{\delta}$. Then we have

$$\tau \sqrt{m} \delta \chi_{\Omega_\kappa} < \sqrt{\sum_{j=k_{\mathrm{dp}}^{\delta}(m)}^{m} (y^\delta, u_j)^2}$$

$$\leq \sqrt{\sum_{j=k_{\mathrm{dp}}^{\delta}(m)}^{m} (y^\delta - y^\dagger, u_j)^2} \chi_{\Omega_\kappa} + \sqrt{\sum_{j=k_{\mathrm{dp}}^{\delta}(m)}^{m} (y^\dagger, u_j)^2}$$

$$\leq \frac{\tau+1}{2} \sqrt{m} \delta + \sigma_{k_{\mathrm{dp}}^{\delta}(m)} \sqrt{\sum_{j=k_{\mathrm{dp}}^{\delta}(m)}^{\infty} (x^\dagger, v_j)^2}$$

$$\implies \frac{\sqrt{m}}{\sigma_{k_{\mathrm{dp}}^{\delta}(m)}} \chi_{\Omega_\kappa} \leq \frac{2}{\tau-1} \sqrt{\sum_{j=k_{\mathrm{dp}}^{\delta}(m)}^{\infty} (x^\dagger, v_j)^2}.$$

Since obviously $k_{\mathrm{dp}}^\delta(m) \le m$, we conclude that

$$(4.8) \qquad \frac{\sqrt{k_{\mathrm{dp}}^\delta}}{\sigma_{k_{\mathrm{dp}}^\delta}} \chi_{\Omega_\kappa} \le \frac{2}{\tau - 1} \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta}^{\infty} (x^\dagger, v_j)^2},$$

whenever $k_{\mathrm{dp}}^\delta \ge k_{\mathrm{st}}^\delta$. Then we have $k_{\mathrm{dp}}^\delta \chi_{\Omega_\kappa} \ge \kappa \chi_{\Omega_\kappa} > \frac{1}{3}\kappa \chi_{\Omega_\kappa}$ and by (4.8) it follows that

$$\sqrt{\sum_{j=1}^{k_{\mathrm{dp}}^\delta} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2}} \chi_{\Omega_\kappa} \le \frac{1}{\sigma_{k_{\mathrm{dp}}^\delta}} \sqrt{\sum_{j=1}^{k_{\mathrm{dp}}^\delta} (y^\delta - y^\dagger, u_j)^2} \chi_{\Omega_\kappa} \le \frac{\tau+1}{2} \frac{\delta\sqrt{k_{\mathrm{dp}}^\delta}}{\sigma_{k_{\mathrm{dp}}^\delta}} \chi_{\Omega_\kappa}$$

$$\le \frac{\tau+1}{\tau-1} \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta}^{\infty} (x^\dagger, v_j)^2}.$$

This proves the assertion (4.7). We come to the approximation error. It holds that

$$\sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{pr}}^\delta-1} (x^\dagger, v_j)^2} \chi_{\Omega_\kappa}$$

$$\le \frac{1}{\sigma_{k_{\mathrm{pr}}^\delta-1}} \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{pr}}^\delta-1} (y^\dagger, u_j)^2} \chi_{\Omega_\kappa}$$

$$\le \frac{1}{\sigma_{k_{\mathrm{pr}}^\delta-1}} \left(\sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{pr}}^\delta-1} (y^\delta, u_j)^2} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{pr}}^\delta-1} (y^\delta - y^\dagger, u_j)^2} \right) \chi_{\Omega_\kappa}$$

$$\le \frac{3\tau+1}{2} \frac{\delta\sqrt{k_{\mathrm{pr}}^\delta-1}}{\sigma_{k_{\mathrm{pr}}^\delta-1}} \chi_{\Omega_\kappa} \le \sqrt{\frac{3}{8}}(3\tau+1) \frac{\sqrt{\sum_{j=1}^{k_{\mathrm{pr}}^\delta-1}(y^\delta - y^\dagger, u_j)^2}}{\sigma_{k_{\mathrm{pr}}^\delta-1}}$$

$$\le \sqrt{\frac{3}{8}}(3\tau+1) \frac{\sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{\infty}(y^\dagger, u_j)^2}}{\sigma_{k_{\mathrm{pr}}^\delta-1}} \le \sqrt{\frac{3}{8}}(3\tau+1) \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{\infty} (x^\dagger, v_j)^2},$$

and consequently

$$\sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \chi_{\Omega_\kappa}$$

$$\le \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{pr}}^\delta-1} (x^\dagger, v_j)^2} \chi_{\Omega_\kappa} + \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{\infty} (x^\dagger, v_j)^2}$$

$$\le \sqrt{\frac{3}{8}}(3\tau+1) \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{\infty} (x^\dagger, v_j)^2}$$

$$(4.9) \qquad \le \sqrt{\frac{3}{8}}(3\tau+1) \left(\sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{k_{\mathrm{st}}^\delta} (x^\dagger, v_j)^2} + \sqrt{\sum_{j=k_{\mathrm{st}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right).$$

We put both estimations together to finish the proof. Note that it is easy to see from Proposition 2.1 that

$$\min_{k\in\mathbb{N}_0} \|x_k^\delta - x^\dagger\|^2 \ge \frac{1}{2} \max\left(\sum_{j=1}^{k_{\mathrm{st}}^\delta-1} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2}, \sum_{j=k_{\mathrm{st}}^\delta+1}^{\infty} (x^\dagger, v_j)^2 \right).$$

We treat the cases $k_{\mathrm{dp}}^\delta < k_{\mathrm{st}}^\delta$ and $k_{\mathrm{dp}}^\delta \geq k_{\mathrm{st}}^\delta$ separately. First, by (4.7),

$$
\begin{aligned}
&\|x_{k_{\mathrm{dp}}^\delta}^\delta - x^\dagger\| \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta \geq k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\sqrt{\sum_{j=1}^{k_{\mathrm{dp}}^\delta} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2}} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta \geq k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\frac{\tau+1}{\tau-1} \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta}^{\infty} (x^\dagger, v_j)^2} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta \geq k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\frac{\tau+1}{\tau-1} + 1 \right) \sqrt{\sum_{j=k_{\mathrm{st}}^\delta}^{\infty} (x^\dagger, v_j)^2} \\
&\leq \left(\frac{\tau+1}{\tau-1} + 1 \right) \left(\sqrt{(x^\dagger, v_{k_{\mathrm{st}}^\delta})^2} + \sqrt{\sum_{j=k_{\mathrm{st}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \\
&\leq \sqrt{2} \left(\frac{\tau+1}{\tau-1} + 1 \right) \left(\min_{k \in \mathbb{N}_0} \|x_k^\delta - x^\dagger\| + \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{k_{\mathrm{st}}^\delta} (x^\dagger, v_j)^2} \right).
\end{aligned}
$$

Second, by (4.9),

$$
\begin{aligned}
&\|x_{k_{\mathrm{dp}}^\delta}^\delta - x^\dagger\| \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta < k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\sqrt{\sum_{j=1}^{k_{\mathrm{dp}}^\delta} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2}} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta < k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\sqrt{\sum_{j=1}^{k_{\mathrm{st}}^\delta-1} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2}} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \chi_{\Omega_\kappa} \\
&\leq \sqrt{2} \min_{k \in \mathbb{N}_0} \|x_k^\delta - x^\dagger\| + \sqrt{\frac{3}{8}} (3\tau+1) \left(\sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{k_{\mathrm{st}}^\delta} (x^\dagger, v_j)^2} + \sqrt{\sum_{j=k_{\mathrm{st}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \\
&\leq \sqrt{2} \left(1 + \sqrt{\frac{3}{8}} (3\tau+1) \right) \left(\min_{k \in \mathbb{N}_0} \|x_k^\delta - x^\dagger\| + \sqrt{\sum_{j=k_{\mathrm{pr}}^\delta}^{k_{\mathrm{st}}^\delta} (x^\dagger, v_j)^2} \right).
\end{aligned}
$$

Thus the proof of Theorem 2.4 is finished with

$$
C_\tau := \sqrt{2} \max \left(\frac{\tau+1}{\tau-1} + 1, 1 + \sqrt{\frac{3}{8}} (3\tau+1) \right). \qquad \square
$$

Proof of Corollary 1.1. This time we treat the cases $k_{\mathrm{dp}}^\delta \leq k_{\mathrm{st}}^\delta$ and $k_{\mathrm{dp}}^\delta > k_{\mathrm{st}}^\delta$ separately. In the latter case, we obtain with (4.7) that

$$
\begin{aligned}
&\|x_{k_{\mathrm{dp}}^\delta}^\delta - x^\dagger\| \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta > k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\frac{\tau+1}{\tau-1} \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta}^{\infty} (x^\dagger, v_j)^2} + \sqrt{\sum_{j=k_{\mathrm{dp}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \right) \chi_{\Omega_\kappa} \chi_{\{k_{\mathrm{dp}}^\delta > k_{\mathrm{st}}^\delta\}} \\
&\leq \left(\frac{\tau+1}{\tau-1} + 1 \right) \sqrt{\sum_{j=k_{\mathrm{st}}^\delta+1}^{\infty} (x^\dagger, v_j)^2} \leq \sqrt{2} \left(\frac{\tau+1}{\tau-1} + 1 \right) \min_{k \in \mathbb{N}_0} \|x_k^\delta - x^\dagger\|.
\end{aligned}
$$

For the former case, note that

$$
\sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{st}}^\delta}(x^\dagger,v_j)^2}\,\chi_{\Omega_\kappa}
$$

$$
\leq \frac{1}{\sigma_{k_{\mathrm{st}}^\delta}}\left(\sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^\delta(k_{\mathrm{st}}^\delta)+1}^{k_{\mathrm{st}}^\delta}(y^\delta,u_j)^2}+\sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{st}}^\delta}(y^\delta-y^\dagger,u_j)^2}\right)\chi_{\Omega_\kappa}
$$

$$
\leq \frac{1}{\sigma_{k_{\mathrm{st}}^\delta}}\left(\tau\sqrt{k_{\mathrm{st}}^\delta}\delta+\frac{\tau+1}{2}\sqrt{k_{\mathrm{st}}^\delta}\delta\right)\leq \frac{3\tau+1}{2}\frac{\sqrt{k_{\mathrm{st}}^\delta}}{\sigma_{k_{\mathrm{st}}^\delta}}\delta.
$$

Then,

$$
\|x_{k_{\mathrm{dp}}^\delta}^\delta-x^\dagger\|\chi_{\Omega_\kappa}\chi_{\{k_{\mathrm{dp}}^\delta\leq k_{\mathrm{st}}^\delta\}}
$$

$$
\leq \sqrt{\sum\nolimits_{j=1}^{k_{\mathrm{st}}^\delta}\frac{(y^\delta-y^\dagger,u_j)^2}{\sigma_j^2}}\,\chi_{\Omega_\kappa}+\sqrt{\sum\nolimits_{j=k_{\mathrm{dp}}^\delta+1}^{k_{\mathrm{st}}^\delta}(x^\dagger,v_j)^2}\,\chi_{\Omega_\kappa}+\sqrt{\sum\nolimits_{j=k_{\mathrm{st}}^\delta+1}^{\infty}(x^\dagger,v_j)^2}
$$

$$
\leq \frac{\tau+1}{2}\frac{\sqrt{k_{\mathrm{st}}^\delta}}{\sigma_{k_{\mathrm{st}}^\delta}}\delta+\frac{3\tau+1}{2}\frac{\sqrt{k_{\mathrm{st}}^\delta}}{\sigma_{k_{\mathrm{st}}^\delta}}\delta+\sqrt{2}\min_{k\in\mathbb{N}_0}\|x_k^\delta-x^\dagger\|
$$

$$
= (2\tau+1)\frac{\sqrt{k_{\mathrm{st}}^\delta}}{\sigma_{k_{\mathrm{st}}^\delta}}\delta+\sqrt{2}\min_{k\in\mathbb{N}_0}\|x_k^\delta-x^\dagger\|.
$$

All we need to do to finish the proof is to control the strong oracle k_{st}^δ. We will need the following two estimates for $k\in\mathbb{N}$ and $p>0$:

$$
\sum_{j=1}^{k}j^p\leq\int_1^{k+1}(x+1)^p\,\mathrm{d}x\leq\frac{(k+2)^{1+p}}{1+p},
$$

$$
\sum_{j=1}^{k}j^p\geq\int_1^{k+1}(x-1)^p\,\mathrm{d}x\geq\frac{k^{1+p}}{1+p}.
$$

For any $k\geq 1$ it holds that

$$
\mathbb{P}\left(\left|\sum_{j=1}^{k}\frac{(y^\delta-y^\dagger,u_j)^2}{\sigma_j^2}-\delta^2\sum_{j=1}^{k}\frac{1}{\sigma_j^2}\right|\geq\frac{\delta^2}{2}\sum_{j=1}^{k}\frac{1}{\sigma_j^2}\right)
$$

$$
\leq \frac{4\mathbb{E}\left[\left|\sum_{j=1}^{k}((y^\delta-y^\dagger,u_j)^2-\delta^2)/\sigma_j^2\right|^2\right]}{\delta^4\left(\sum_{j=1}^{k}1/\sigma_j^2\right)^2}=\frac{4\delta^4\sum_{j=1}^{k}\mathbb{E}[|(Z,u_j)^2-1|^2/\sigma_j^4]}{\delta^4\left(\sum_{j=1}^{k}1/\sigma_j^2\right)^2}
$$

$$
\leq \frac{8(\gamma_4+1)\sum_{j=1}^{k}1/\sigma_j^4}{\left(\sum_{j=1}^{k}1/\sigma_j^2\right)^2}
$$

$$
\leq \frac{(8(\gamma_4+1)/c_q^2)\sum_{j=1}^{k}j^{2q}}{(1/C_q^2)\left(\sum_{j=1}^{k}j^q\right)^2}
$$

$$
\leq \frac{8(\gamma_4+1)C_q^2}{c_q^2}\frac{(1/(2q+1))(k+2)^{2q+1}}{((1/(q+1))k^{q+1})^2}
$$

$$
\tag{4.10}\leq \frac{8(\gamma_4+1)C_q^2(q+1)^2 3^{2q+1}}{c_q^2(2q+1)}\frac{1}{k}=\frac{(\gamma_4+1)\tilde{C}_q}{k},
$$

with

$$\tilde{C}_q := \frac{8C_q^2(q+1)^2 3^{2q+1}}{c_q^2(2q+1)}.$$

Similar to $\bar{k}_{\text{pr}}^\delta$ we define

$$\bar{k}_{\text{st}}^\delta := \min\left\{ k \geq 0 : \delta^2 \sum_{j=1}^k \frac{1}{\sigma_j^2} \geq \sum_{j=k+1}^\infty (x^\dagger, v_j)^2 \right\}$$

and set

$$\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta} := \left\{ \left| \sum_{j=1}^{\lceil \bar{k}_{\text{st}}^\delta/3 \rceil} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} - \delta^2 \sum_{j=1}^{\lceil \bar{k}_{\text{st}}^\delta/3 \rceil} \frac{1}{\sigma_j^2} \right| \geq \frac{\delta^2}{2} \sum_{j=1}^{\lceil \bar{k}_{\text{st}}^\delta/3 \rceil} \frac{1}{\sigma_j^2} \right\}$$

$$\cap \left\{ \left| \sum_{j=1}^{3\bar{k}_{\text{st}}^\delta} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} - \delta^2 \sum_{j=1}^{3\bar{k}_{\text{st}}^\delta} \frac{1}{\sigma_j^2} \right| \geq \frac{\delta^2}{2} \sum_{j=1}^{3\bar{k}_{\text{st}}^\delta} \frac{1}{\sigma_j^2} \right\}.$$

Note that $\bar{k}_{\text{st}}^\delta \geq \bar{k}_{\text{pr}}^\delta$; cf. Proposition 2.3. On $\chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}}$ it holds that

(4.11)
$$\frac{\bar{k}_{\text{st}}^\delta}{3} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq k_{\text{st}}^\delta \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} - 1 < k_{\text{st}}^\delta \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq 3\bar{k}_{\text{st}}^\delta \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}}.$$

We show the claim (4.11). It holds that

$$\sum_{j=1}^{\lceil \bar{k}_{\text{st}}^\delta/3 \rceil} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq \frac{3\delta^2}{2} \sum_{j=1}^{\lceil \bar{k}_{\text{st}}^\delta/3 \rceil} \frac{1}{\sigma_j^2} \leq \frac{3\delta^2}{2} \left(\frac{\lceil \bar{k}_{\text{st}}^\delta/3 \rceil}{\bar{k}_{\text{st}}^\delta - 1} \sum_{j=1}^{\bar{k}_{\text{st}}^\delta - 1} \frac{1}{\sigma_j^2} \right)$$

$$\leq \delta^2 \sum_{j=1}^{\bar{k}_{\text{st}}^\delta - 1} \frac{1}{\sigma_j^2} < \sum_{j=\bar{k}_{\text{st}}^\delta}^\infty (x^\dagger, v_j)^2 \leq \sum_{j=\lceil \bar{k}_{\text{st}}^\delta/3 \rceil + 1}^\infty (x^\dagger, v_j)^2,$$

where we have used that σ_j is monotonically decreasing in the second step, and that $\bar{k}_{\text{st}}^\delta \geq \kappa \geq 3$ in the third and in the last step. This shows that $(\bar{k}_{\text{st}}^\delta/3)\chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq k_{\text{st}}^\delta - 1$. For the second claim we can assume without loss of generality that $\sum_{j=3\bar{k}_{\text{st}}^\delta+1}^\infty (x^\dagger, v_j)^2 > 0$. Then we have

$$\sum_{j=1}^{3\bar{k}_{\text{st}}^\delta} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \geq \frac{\delta^2}{2} \sum_{j=1}^{3\bar{k}_{\text{st}}^\delta} \frac{1}{\sigma_j^2} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \geq \frac{3\delta^2}{2} \sum_{j=1}^{\bar{k}_{\text{st}}^\delta} \frac{1}{\sigma_j^2}$$

$$\geq \frac{3}{2} \sum_{j=\bar{k}_{\text{st}}^\delta+1}^\infty (x^\dagger, v_j)^2 > \sum_{j=3\bar{k}_{\text{st}}^\delta+1}^\infty (x^\dagger, v_j)^2,$$

and we deduce that $k_{\text{st}}^\delta \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq 3\bar{k}_{\text{st}}^\delta$. Now we are ready to finish the proof of Corollary 1.1, as follows:

$$\delta^2 \frac{k_{\text{st}}^\delta}{\sigma_{k_{\text{st}}^\delta}^2} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}}$$

$$\leq \delta^2 \frac{3\bar{k}_{\text{st}}^\delta}{\sigma_{3\bar{k}_{\text{st}}^\delta}^2} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq \frac{\delta^2}{c_q} (3\bar{k}_{\text{st}}^\delta)^{1+q} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}} \leq \frac{9^{1+q}\delta^2}{c_q} \left\lceil \frac{\bar{k}_{\text{st}}^\delta}{3} \right\rceil^{1+q} \chi_{\tilde{\Omega}_{\bar{k}_{\text{st}}^\delta}}$$

$$\leq \frac{9^{1+q}\delta^2(1+q)C_q}{c_q} \sum_{j=1}^{\lceil \bar{k}_{\mathrm{st}}^\delta/3 \rceil} \frac{1}{\sigma_j^2} \chi_{\tilde{\Omega}_{\bar{k}_{\mathrm{st}}^\delta}} \leq \frac{9^{1+q}(1+q)C_q}{c_q} 2 \sum_{j=1}^{\lceil \bar{k}_{\mathrm{st}}^\delta/3 \rceil} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} \chi_{\tilde{\Omega}_{\bar{k}_{\mathrm{st}}^\delta}}$$

$$\leq \frac{9^{1+q}(1+q)C_q}{c_q} 2 \sum_{j=1}^{k_{\mathrm{st}}^\delta - 1} \frac{(y^\delta - y^\dagger, u_j)^2}{\sigma_j^2} \leq \frac{9^{1+q}(1+q)C_q}{c_q} 4 \min_{k \in \mathbb{N}_0} \|x_k^\delta - x^\dagger\|^2.$$

Putting all this together, we obtain

$$\|x_{k_{\mathrm{dp}}^\delta}^\delta - x^\dagger\| \chi_{\tilde{\Omega}_{\bar{k}_{\mathrm{st}}^\delta}} \chi_{\Omega_\kappa}$$

$$\leq \max\left(\sqrt{2}\left(\frac{\tau+1}{\tau-1}+1\right), \sqrt{2} + (2\tau+1)\sqrt{\frac{9^{1+q}(1+q)C_q}{c_q}4} \right) \min_{k \in \mathbb{N}_0} \|x_k^\delta - x^\dagger\|,$$

and the proof follows with

$$\mathbb{P}\big(\chi_{\Omega_\kappa}\chi_{\tilde{\Omega}_{\bar{k}_{\mathrm{st}}^\delta}}\big) \geq 1 - \mathbb{P}(\chi_{\Omega_\kappa}) - \mathbb{P}\big(\chi_{\tilde{\Omega}_{\bar{k}_{\mathrm{st}}^\delta}}\big)$$

$$\geq 1 - \max\left(\frac{12}{\tau^2 + 2\tau - 3}, 9\right) \frac{\sqrt{1+\gamma_4}}{\sqrt{\kappa}} - \frac{2(\gamma_4+1)\bar{C}_q}{\kappa},$$

where we used (4.2) with Proposition 2.5 and (4.10). $\qquad\square$

5. Numerical comparison. In this section we compare k_{dp}^δ with k_{bal}^δ and k_{es}^δ numerically. As examples, we take four model problems from the open-source MATLAB toolbox "Regutools" [13], namely, phillips and deriv2 (mildly ill-posed), on the one hand, and gravity and heat (severely ill-posed), on the other. They cover various settings and are discretisations of the Volterra/Fredholm integral equations, which are solved by means of either quadrature rules or Galerkin methods. Note that the problems behind deriv2 (the second derivative) and heat (the backwards heat equation) are also part of the recent numerical survey [37]. The discretisation dimension for the examples is fixed at $D = 5000$. As measurement noise, we choose Gaussian white noise, and the singular value decomposition is determined numerically with the function csvd from the toolbox. The simulations are run for noise levels $\delta = 10^0$, 10^{-2}, 10^{-4} and 10^{-6}. The statistical quantities are computed by 1000 independent Monte Carlo samples. Similar to the numerical survey [37], we set $\tau = 1.5$ in the definition of the discrepancy principle, and $\kappa = 4$ in the definition of the balancing principle, i.e., for the balancing principle we choose a δ-independent fudge parameter. Note that for $\kappa = 4$ the probability p_κ of a bad event from Example 3.2 is extremely small compared to that for only 100 consecutive runs.

For comparison, we also calculated the (clearly inaccessible) optimal truncation level $k_{\mathrm{opt}}^\delta := \arg\min_{k \in \mathbb{N}} \|x_k^\delta - x^\dagger\|$. We express the sample mean of $e_* := \|x_{k_*^\delta}^\delta - x^\dagger\|$ (where $* = $ dp, bal, es, opt) together with the estimated standard deviation in tabular form in Tables 5.1–5.4. In Tables 5.5–5.8 we also present the statistics of the corresponding truncation levels k_*^δ together with the optimal stopping time $k_{\mathrm{opt}}^\delta = \arg\min \|x_k^\delta - x^\dagger\|$ and the weak and strong oracles k_{pr}^δ and k_{st}^δ. It can be clearly seen that k_{dp}^δ and k_{opt}^δ are close to k_{pr}^δ and k_{st}^δ, respectively, in accordance with Proposition 2.1. Interestingly, k_{bal}^δ tends to be close to k_{dp}^δ and hence k_{pr}^δ instead of k_{st}^δ. This is probably due to the comparatively large fudge parameter $\kappa = 4$. Choosing a substantially smaller fudge parameter κ, however, is less stable, and we decided to stick to the choice from [37]. We also see that the error for k_{es}^δ is quite large. From Tables 5.5–5.8 it is not clear if this is due to very rare events, where one substantially stops too late, or if there is a regular late stopping. We visualise the statistics of k_{pr}^δ as boxplots in

Figure 5.1. While the median is relatively close to k_{pr}^{δ}, in a substantial part of all runs it stops way too late. In one quarter of all runs, the stopping happens later than the corresponding index of the upper border of the blue box. However, k_{es}^{δ} concentrates more and more as the noise level decreases. This result indicates that the early stopping discrepancy principle is better suited for comparably smaller noise levels. All in all the numerical experiments show that the choice $\tau = 1$ for the early stopping discrepancy principle is probably too aggressive for general linear problems violating the assumptions in [5]. Note that we observed in simulations that adding a small fudge parameter in (3.3), e.g., $\tau = 1.01$, increases stability a lot for the early stopping discrepancy principle and yields an error slightly larger than, but comparable to, the other two methods, at least for the mildly ill-posed examples under small error levels. Increasing the fudge parameter further improves stability, but at the cost of less accuracy.

TABLE 5.1
Sample mean and standard deviation of e_* for `phillips`.

δ	e_{dp}	$\mathrm{std}(e_{\mathrm{dp}})$	e_{bal}	$\mathrm{std}(e_{\mathrm{bal}})$	e_{es}	$\mathrm{std}(e_{\mathrm{es}})$	e_{opt}	$\mathrm{std}(e_{\mathrm{opt}})$
e0	1.0e0	(1e-1)	1.1e0	(2e-1)	8.9e4	(4e5)	6.3e-1	(2e-1)
e-2	7.6e-2	(1e-3)	7.6e-2	(1e-3)	5.7e2	(2e3)	6.7e-2	(1e-2)
e-4	1.3e-2	(3e-4)	1.3e-2	(3e-4)	5.2e0	(1e1)	1.1e-2	(2e-3)
e-6	2.6e-3	(3e-4)	2.9e-3	(1e-4)	1.1e-1	(3e-1)	1.7e-3	(2e-3)

TABLE 5.2
Sample mean and standard deviation of e_* for `deriv2`.

δ	e_{dp}	$\mathrm{std}(e_{\mathrm{dp}})$	e_{bal}	$\mathrm{std}(e_{\mathrm{bal}})$	e_{es}	$\mathrm{std}(e_{\mathrm{es}})$	e_{opt}	$\mathrm{std}(e_{\mathrm{opt}})$
e0	8.3e0	(3e1)	1.8e0	(0)	4.7e5	(1e6)	1.7e0	(2e-1)
e-2	1.0e0	(4e-1)	9.5e-1	(8e-3)	3.2e3	(9e3)	8.5e-1	(8e-2)
e-4	4.7e-1	(1e-2)	5.2e-1	(3e-2)	2.7e1	(6.0e1)	4.1e-1	(2e-2)
e-6	2.3e-1	(4e-3)	2.5e-1	(4e-3)	5.7e-1	(1e0)	2.0e-1	(5e-3)

TABLE 5.3
Sample mean and standard deviation of e_* for `gravity`.

δ	e_{dp}	$\mathrm{std}(e_{\mathrm{dp}})$	e_{bal}	$\mathrm{std}(e_{\mathrm{bal}})$	e_{es}	$\mathrm{std}(e_{\mathrm{es}})$	e_{opt}	$\mathrm{std}(e_{\mathrm{opt}})$
e0	4.6e0	(6e-1)	5.0e0	(8e-2)	5.2e15	(9e15)	3.2e0	(6e-1)
e-2	8.9e-1	(2e-1)	9.5e-1	(2e-1)	4.0e13	(7e13)	6.5.8e-1	(1e-1)
e-4	1.8e-1	(2e-3)	1.8e-1	(7e-3)	5.2e11	(9e11)	1.2e-1	(2e-2)
e-6	4.3e-2	(1e-3)	3.9e-2	(6e-3)	3.8e9	(8e9)	2.5e-2	(4e-3)

TABLE 5.4
Sample mean and standard deviation of e_* for `heat`.

δ	e_{dp}	$std(e_{dp})$	e_{bal}	$std(e_{bal})$	e_{es}	$std(e_{es})$	e_{opt}	$std(e_{opt})$
e0	1.7e1	(6e0)	1.7e1	(5e-1)	5.6e7	(4e8)	1.4e1	(1e0)
e-2	4.8e0	(3e-1)	5.2e0	(5e-2)	1.1e5	(5e5)	3.1e0	(5e-1)
e-4	6.4e-1	(7e-2)	7.5e-1	(6e-2)	1.1e2	(4e2)	3.8e-1	(3e-2)
e-6	1.7e-1	(3e-3)	1.8e-1	(5e-3)	8.4e1	(6e2)	1.1e-1	(8e-3)

TABLE 5.5
Sample mean and standard deviation of k_* for `phillips`.

δ	k_{dp}	$std(k_{dp})$	k_{bal}	$std(k_{bal})$	k_{es}	$std(k_{es})$	k_{opt}	$std(k_{opt})$	k_{pr}	$std(k_{pr})$	k_{st}	$std(k_{st})$
e0	3.3	(0.7)	2.9	(0.5)	47.5	(63.7)	5	(0.6)	4.5	(0.8)	5.2	(0.5)
e-2	7	(0)	7	(0)	43.7	(53.8)	8.1	(1.2)	7	(0,1)	9.4	(0.7)
e-4	12	(0)	12	(0)	48.5	(51.2)	15.9	(2)	12.1	(0.6)	17.4	(1)
e-6	26.1	(1)	25.1	(0.4)	59.7	(60.3)	34.3	(3.3)	28	(1)	35.3	(1.2)

TABLE 5.6
Sample mean and standard deviation of k_* for `deriv2`.

δ	k_{dp}	$std(k_{dp})$	k_{bal}	$std(k_{bal})$	k_{es}	$std(k_{es})$	k_{opt}	$std(k_{opt})$	k_{pr}	$std(k_{pr})$	k_{st}	$std(k_{st})$
e0	0.3	(0.5)	0	(0)	43.5	(72)	1.1	(0.3)	1	(0)	1.1	(0.3)
e-2	1.3	(0.6)	1	(0)	39	(62)	3.2	(0.8)	2	(0.7)	3.4	(0.8)
e-4	7	(0.5)	46	(54)	48.6	(59.5)	12.4	(1.7)	8.7	(0.8)	14.6	(1.5)
e-6	32	(1.1)	27	(1)	53	(38)	53	(4.6)	39	(1.5)	66.5	(3.2)

TABLE 5.7
Sample mean and standard deviation of k_* for `gravity`.

δ	k_{dp}	$std(k_{dp})$	k_{bal}	$std(k_{bal})$	k_{es}	$std(k_{es})$	k_{opt}	$std(k_{opt})$	k_{pr}	$std(k_{pr})$	k_{st}	$std(k_{st})$
e0	4.3	(0.5)	4	(0.1)	44.7	(61)	5.9	(0.8)	4.2	(0.4)	5.7	(0.6)
e-2	8.7	(0.5)	8.5	(0.5)	28.2	(47.1)	10.5	(0.8)	8.2	(0.4)	10.3	(0.6)
e-4	13	(0.1)	13	(0.2)	48.6	(59.5)	14.9	(0.7)	12.7	(0.4)	14.7	(0.6)
e-6	17	(0.1)	17.4	(0.5)	43.7	(49.4)	19.3	(0.7)	17	(0)	19.2	(0.6)

TABLE 5.8
Sample mean and standard deviation of k_* for `heat`.

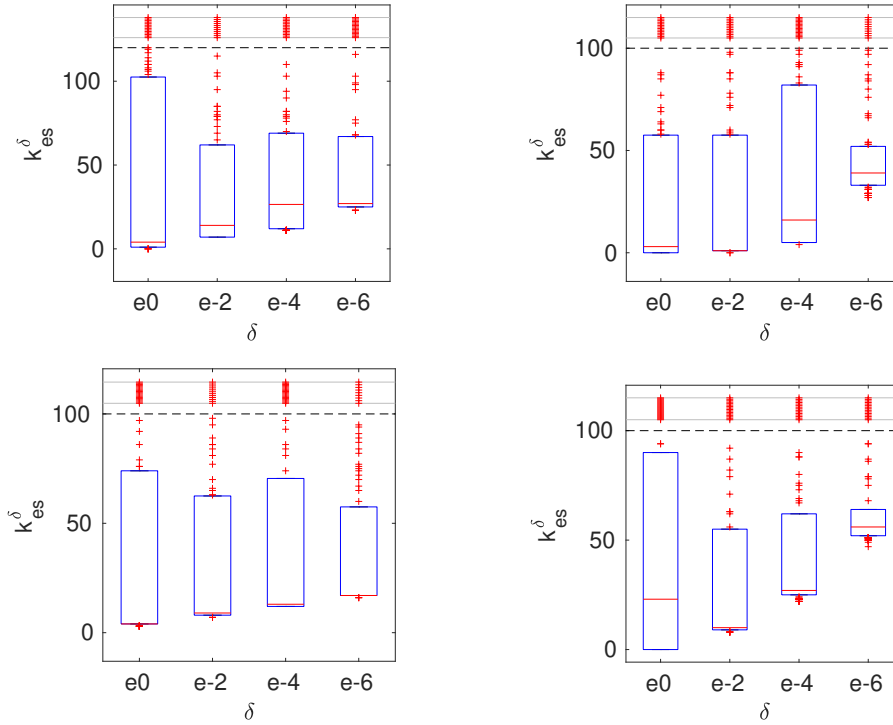| δ	k_{dp}	$std(k_{dp})$	k_{bal}	$std(k_{bal})$	k_{es}	$std(k_{es})$	k_{opt}	$std(k_{opt})$	k_{pr}	$std(k_{pr})$	k_{st}	$std(k_{st})$
e0	1.5	(0.9)	0.1	(0.3)	53	(66)	3.3	(0.8)	1.6	(0.6)	2.9	(0.9)
e-2	9.9	(0.8)	9	(0.1)	41	(53)	17	(1)	9.4	(0.5)	13.8	(0.8)
e-4	26	(0.6)	25	(1.4)	48	(38)	32	(2)	24.1	(1.3)	29.4	(1)
e-6	51	(0.4)	50	(0.5)	72	(42)	67	(2)	49.4	(0.9)	62.4	(1.5)

FIG. 5.1. *Boxplots of the 100 realisations of k_{es}^{δ} for decreasing noise level δ and the different test problems: upper left,* `phillips`*; upper right,* `deriv2`*; lower left,* `gravity`*; and lower right,* `heat`*. On each blue box, the red bar indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Red crosses depict outliers, which are data points falling outside the blue box.*

REFERENCES

[1]   F. BAUER AND T. HOHAGE, *A Lepskij-type stopping rule for regularized Newton methods*, Inverse Problems, 21 (2005), pp. 1975–1991.

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

[3]   F. BAUER AND M. REISS, *Regularization independent of the noise level: an analysis of quasi-optimality*, Inverse Problems, 24 (2008), Art. 055009, 16 pages.

[4]   N. BISSANTZ, T. HOHAGE, A. MUNK, AND F. RUYMGAART, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM J. Numer. Anal., 45 (2007), pp. 2610–2636.

[5]   G. BLANCHARD, M. HOFFMANN, AND M. REISS, *Early stopping for statistical inverse problems via truncated SVD estimation*, Electron. J. Stat., 12 (2018), pp. 3204–3231.

[6]   ———, *Optimal adaptation for early stopping in statistical inverse problems*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 1043–1075.

[7]   G. BLANCHARD AND P. MATHÉ, *Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration*, Inverse Problems, 28 (2012), Art. 115011, 23 pages.

[8]   L. CAVALIER, *Inverse problems in statistics*, in Inverse Problems and High-Dimensional Estimation, P. Alquier, E. Gautier, and G. Stoltz, eds., vol. 203 of Lect. Notes Stat. Proc., Springer, Heidelberg, 2011, pp. 3–96.

[9]   D. L. DONOHO AND I. M. JOHNSTONE, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[10] A. GOLDENSHLUGER, *On pointwise adaptive nonparametric deconvolution*, Bernoulli, 5 (1999), pp. 907–925.

[11] A. GUT, *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[12] P. C. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.

[13] ———, *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.

[14] B. HARRACH, T. JAHN, AND R. POTTHAST, *Beyond the Bakushinkii veto: regularising linear inverse problems without knowing the noise distribution*, Numer. Math., 145 (2020), pp. 581–603.

[15] ———, *Regularizing linear inverse problems under unknown non-Gaussian white noise allowing repeated measurements*, IMA J. Numer. Anal., published online 2022. https://doi.org/10.1093/imanum/drab098

[16] K. ITO AND B. JIN, *Regularized linear inversion with randomized singular value decomposition*, in Mathematical and Numerical Approaches for Multi-Wave Inverse Problems, L. Beilina, M. Bergounioux, M. Cristofol, A. Da Silva, and A. Litman, eds., vol. 328 of Springer Proc. Math. Stat., Springer, Cham, 2020, pp. 45–72.

[17] T. JAHN, *Optimal convergence of the discrepancy principle for polynomially and exponentially ill-posed operators under white noise*, Numer. Funct. Anal. Optim., (43) (2022), pp. 145–167.

[18] S. KINDERMANN, S. PEREVERZYEV, JR., AND A. PILIPENKO, *The quasi-optimality criterion in the linear functional strategy*, Inverse Problems, 34 (2018), Art. 075001, 24 pages.

[19] M. LAVRENTIEV, *O Nekotorykh Nekorrektnykh Zadachakh Matematicheskoi Fiziki (On Some Ill-Posed Problems of Mathematical Physics)*, Sb. USSR. Acad., Novosibirsk, 1962.

[20] O. LEPSKII, *On a problem of adaptive estimation in Gaussian white noise*, Theory Probab. Appl., 35 (1991), pp. 454–466.

[21] H. LI AND F. WERNER, *Empirical risk minimization as parameter choice rule for general linear regularization methods*, Ann. Inst. Henri Poincaré Probab. Stat., 56 (2020), pp. 405–427.

[22] S. LU AND P. MATHÉ, *Discrepancy based model selection in statistical inverse problems*, J. Complexity, 30 (2014), pp. 290–308.

[23] F. LUCKA, K. PROKSCH, C. BRUNE, N. BISSANTZ, M. BURGER, H. DETTE, AND F. WÜBBELING, *Risk estimators for choosing regularization parameters in ill-posed problems—properties and limitations*, Inverse Probl. Imaging, 12 (2018), pp. 1121–1155.

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

[25] P. MATHÉ, *The Lepskiĭ principle revisited*, Inverse Problems, 22 (2006), pp. L11–L15.

[26] P. MATHÉ AND B. HOFMANN, *How general are general source conditions?*, Inverse Problems, 24 (2008), Art. 015009, 5 pages.

[27] P. MATHÉ AND S. V. PEREVERZEV, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), pp. 789–803.

[28] ———, *The discretized discrepancy principle under general source conditions*, J. Complexity, 22 (2006), pp. 371–381.

[29] ———, *Regularization of some linear ill-posed problems with discretized random noisy data*, Math. Comp., 75 (2006), pp. 1913–1929.

[30] ———, *Complexity of linear ill-posed problems in Hilbert space*, J. Complexity, 38 (2017), pp. 50–67.

[31] V. A. MOROZOV, *The principle of disparity in solving operator equations by the method of regularization*, Ž. Vyčisl. Mat i Mat. Fiz., 8 (1968), pp. 295–309.

[32] C. M. STEIN, *Estimation of the mean of a multivariate normal distribution*, Ann. Statist., 9 (1981), pp. 1135–1151.

[33] A. N. TIKHONOV AND V. B. GLASKO, *Use of the regularization method in non-linear problems*, U.S.S.R. Comput. Math. and Math. Phys., 5 (1965), pp. 93–107.

[34] A. TSYBAKOV, *On the best rate of adaptive estimation in some inverse problems*, C. R. Acad. Sci. Paris Sér. I Math., 330 (2000), pp. 835–840.

[35] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.

[36] G. WAHBA, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., 14 (1977), pp. 651–667.

[37] F. WERNER, *Adaptivity and oracle inequalities in linear statistical inverse problems: a (numerical) survey*, in New Trends in Parameter Identification for Mathematical Models, B. Hofmann, A. Leitão, and J. P. Zubelli, eds., Trends Math., Birkhäuser/Springer, Cham, 2018, pp. 291–316.