INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

Master in Mathematic Xavier Loizeau

aus Nantes, France

Tag der mündlichen Prüfung:

Hierarchical Bayes and frequentist aggregation in inverse problems

Betreuer: Prof. Dr. Jan Johannes

Acknowledgments

My first thought obviously goes to my supervisor, Jan Johannes, who gave me the opportunity to follow this PhD program. I have been lucky to receive your guidance, as, asking for your help, I have always been sure to find patience, and attentiveness; your personal support have been primordial.

For their fruitful suggestions, I would also like to thank Fabienne Comte, Claudia Schilings, Björn Sprungk, and all the members of the RTG 1953. I would like to mention in particular the members of the statistic group of Heidelberg University, for the welcoming and studious atmosphere they conveyed.

For their personal support I would like to thank Alexander and Jacob with whom I could always share my doubts and troubles but also learned so much; my boulering friends and in particular Marilena and Tobias who initiated me; mes amis qui, malgré la distance, m'ont soutenu, je pense en particulier à Alexandre, Fabien, Marin, et Marine; mais aussi à Cindy, Lubin, Perrine, Stephane, et Sydney.

Je souhaiterais également remercier ma famille qui m'a toujours porté dans mes études. Merci, donc, à mes parents, et à mes frères et sœur Jacques-Henri, Claire, et Hugues, de m'avoir toujours apporté l'aide, l'environnement, l'attention, et l'affection qui m'ont permis d'arriver là où je suis.

Enfin, je souhaite remercier celles et ceux qui m'ont montré l'exemple lors de mes études scientifiques, Jacqueline Heuzé, Hugues et Claire Loizeau, François Sauvageot, et, encore une fois, Jan Johannes.

I am grateful to the Deutsche Forschungs Gesellschaft (DFG) for their support that was accompanied with the work in the Research Training Group (RTG) 1953 "Statistical Modelling of Complex Systems and Processes" as well as to the administration of the *Faculty of Mathematics and Informatic* of the Ruprecht-Karls-Universität Heidelberg.

Zusammenfassung

In dieser Arbeit betrachten wir lineare, schlecht gestellte, statistische Probleme von zwei Standpunkten aus: Der Bayesschen und der frequentistischen Sichtweise.

In der Bayesschen Betrachtung untersuchen wir zwei verschiedene Methoden zur asymptotischen Analyse von Gaußschen Sieve-Priors und ihren hierarchischen Gegenstücken.
Zuerst untersuchen wir eine iterative Methode, in welcher die a-posteriori Verteilung als a-priori Verteilung verwendet wird um eine neue a-posteriori Verteilung zu berechnen. Die Likelihood und Daten bleiben dabei unverändert. Wir interessieren uns für das asymptotische Verhalten dieses Prozesses (Existenz und Bestimmung der Grenzverteilung).
Im zweiten (klassischen) Ansatz untersuchen wir das Verhalten der a-posteriori Verteilung, wenn die Anzahl der Datenpunkte wächst. Unter der Annahme, dass ein wahrer Parameter existiert, interessieren wir uns dafür, ob sich die a-posteriori Verteilung um diesen Parameter mit einer optimalen Rate konzentriert.
Die Ergebnisse aus beiden Fällen werden auf das inverse Gaußsche Folgenraummodell angewandt.
Schließlich beweisen wir, dass durch den Mittelwert der a-posteriori Verteilung im hierarchischen Gaußschen Sieve-Prior Ansatz sowohl ein Shrinkage- als auch ein Aggregation-Schätzer gegeben ist, welcher interessante Optimalitätseigenschaften hat.

Diese Ergebnisse über den Mittelwert der a-posteriori Verteilung von Gaußschen Sieve-Priors motivieren die Untersuchung des quadratischen Fehlers von Aggregation-Schätzern, deren Form obigen Mittelwerten von a-posteriori Verteilungen ähnelt. Wir stellen (highlight) eine Strategie vor, welche auf einer Zerlegung des Fehlers beruht. Auf diese Weise können wir für einen bekannten bzw. unbekannten Operator und für unabhängige bzw. absolut reguläre Daten optimale Konvergenzraten finden.
Wir illustrieren diese Methode am inversen Gaußschen Folgenraummodell sowie anhand von zyklischer Dekonvolution, indem wir optimale Raten unter schwachen Annahmen beweisen.

Abstract

Considering a family of statistical, linear, ill-posed inverse problems, we propose their study from two perspectives, the Bayesian and frequentist paradigms.

Under the Bayesian paradigm, we investigate two different asymptotic analyses for Gaussian sieve priors and their hierarchical counterpart.
The first analysis is with respect to an iteration procedure, where the posterior distribution is used as a prior to compute a new posterior distribution while using the same likelihood and data. We are interested in the limit of the sequence of distributions generated this way, if it exists.
The second analysis, more traditionally, investigates the behaviour of the posterior distribution as the amount of data increases. Assuming the existence of a true parameter, one is then interested in showing that the posterior distribution contracts around the truth at an optimal rate.
We illustrate all those results by their application to the inverse Gaussian sequence space model.
Finally we exhibit that the posterior mean of the hierarchical Gaussian sieve prior is both a shrinkage and an aggregation estimator, with interesting optimality properties.

Motivated by the last findings about posterior mean of hierarchical Gaussian sieves, we propose to investigate the quadratic risk of aggregation estimators, which shape mimics the one of the above-mentioned posterior means. We introduce a strategy, relying on the decomposition of the risk, which allows to obtain optimal convergence rates in the cases of known and unknown operator, for dependent as well as absolutely regular data. We demonstrate the use of this method on the inverse Gaussian sequence space model as well as the circular density deconvolution and obtained optimality results under mild hypotheses.

Contents

Introduction

As the interests of scientific research become more and more complex, the complexity of measuring quantities related to those interests also increases. In an attempt to rise to this challenge, we design equally complex systems, which output vast amounts of data which are only remotely linked to the phenomenon of interest and fundamentally random. In such a context, it is urging to design statistical methods which are fit to leverage such data. We hence propose here a study of linear statistical ill-posed inverse problems, a family of models which may arise in the framework we just described, and investigate statistical methods for estimation within them.

In chapter 1, we provide a brief overview of the theory of linear statistical ill-posed inverse problems. To do so, we begin, in section 1.1, by presenting a definition of those models and some of their most common hypotheses and difficulties.
As we are interested in the case of random data for inverse problems, we present in section 1.2 the notion of stochastic process, a general formulation for the data when considering the statistical version of linear ill-posed inverse problems. In particular, we will consider four flavours for those; when the data are either independent or when they form an absolutely regular process; and when we suppose that the operator of the inverse problem is known or when we need to observe a second set of data to learn about the operator.
We will consider the study of those models under the two paradigms of Bayesian and frequentist inference. We will hence introduce those two approaches. First the frequentist approach in section 1.3 where we will introduce the notion of estimator as well as notions of decision theory used to quantify the quality of an estimator and to define notions of optimality. We then proceed with the Bayesian approach in section 1.4 where we give a short reminder of necessary conditions for the posterior distribution to exist in a satisfying sense; we then introduce an iteration procedure which allows to define non-informative priors; and we present the way in which we will quantify the quality of posterior distribution thanks to the pragmatic Bayesian approach.
We conclude this overview with the introduction of the two models we will use to illustrate our methods, namely, the inverse Gaussian sequence space in section 1.5, as well as the circular density deconvolution in section 1.6.

While considering the Bayesian paradigm in chapter 2, we propose the study of inverse problems using Gaussian process sieve priors as well as their hierarchical counterpart where the threshold is a random variable.
We investigate two different asymptotic analyses. The first asymptotic faces the difficulty to justify the choice of a particular prior in the non-parametric context, when prior in-

formation can be little reliable. We then study a non informative prior obtained by an iteration of the posterior where a posterior distribution is used as a new prior, used with the same data and likelihood, to compute a new posterior distribution, over and over again. This procedure generates a family of posterior distributions, giving more and more weight to the observations while the prior information fades away. If it exists, the distribution obtained when this iteration parameter tends to infinity is called self informative Bayes carrier while its mean is called self informative limit. We show in section 2.1 that, under a continuity assumption for the likelihood, the self informative Bayes carrier for Gaussian sieves is supported by the set of maximisers of the likelihood. For the hierarchical sieves, we show in section 2.2 that the self informative Bayes carrier of the threshold parameter is supported by the set of minimisers of a penalised contrast, which shows a link between this method and the frequentist estimation via penalised contrast minimisation.
The second asymptotic, more traditionally, investigates the behaviour of the posterior distribution as the quality (or amount) of the data increases. Considering the classical notion of posterior contraction rate and uniform contraction rate, we present in section 2.3 two technics to compute upper bounds for them. The first, presented in section 2.3.1 relies on the computation of the posterior moments of the distance between the true parameter and the random parameter and it allows us to show optimal bound for the Gaussian sieve in the case of the inverse Gaussian sequence space model. The second, presented in section 2.3.2 is specific to the hierarchical Gaussian sieve and relies on a decomposition of the posterior risk. We then proceed to show in section 2.4 that those methods may apply in the context of the inverse Gaussian sequence space model. All those contraction results are obtained for all values of the iteration parameter, including in the limiting case and hence give us a proof for the optimality of the penalised contrast minimisation estimator in terms of convergence in probability.
Finally we exhibit in section 2.5 that the posterior mean of the hierarchical Gaussian sieve prior is both a shrinkage and an aggregation estimator, with interesting optimality properties.

Motivated by the last findings about posterior mean of hierarchical Gaussian sieves, we propose in chapter 3 to investigate the quadratic risk of a family of aggregation estimators, which shape, presented in section 3.1 mimics the one of the above-mentioned posterior means. In section 3.2 we highlight a strategy, relying on the decomposition of the risk, which allows to obtain optimal convergence rates in the cases of known and unknown operator, for dependent as well as absolutely regular data. Finally, we apply this strategy the the inverse Gaussian sequence space model in section 3.3, both in the known operator case (section 3.3.1), and the unknown operator case (section 3.3.2); and to the circular density deconvolution model in section 3.4, in the known operator with independent data case (section 3.4.1), known operator and absolutely regular process data case (section 3.4.2), and to the case of an unknown operator (section 3.4.3).

Chapter 1

Background and review

As stated in the introduction, we propose in this thesis to consider the problem of parameter estimation in the context of statistical ill-posed linear inverse problems under two different paradigms, the frequentist and the Bayesian paradigm respectively. As a consequence, it is suitable to define with precision this family of problem and those two paradigms. That is what we aim to do in the following chapter with the following section structure.

In section 1.1, we give a brief formulation of linear inverse problems and the difficulties that arise as a consequence of their specific structure. We then present in section 1.2 the notion of stochastic process which generalises the type of data we will consider in our examples. We also formulate how the stochastic processes we will observe relate to the parameters of inverse problems as well as the different dependence structures we might consider.

We then move on to consider the frequentist paradigm in section 1.3. At first we consider the notion of estimator and, in particular, a form of estimators that arises naturally with our data. Referring back to the specificities of inverse problems, we highlight the importance of so called regularisation methods, and give particular interest to the family regularisation by dimension reduction which we will use throughout the thesis. Notions of decision theory which let us define what is a satisfying estimator are then presented and we illustrate those notions with their application in our context.

In section 1.4 we consider the Bayesian paradigm . After briefly introducing the keystones of this paradigm, we give some examples of widely used prior distributions for stochastic processes. We will then consider a generalisation of the posterior distribution through an iteration procedure previously introduced in Bunke and Johannes (2005). Underlining the need for a quantification of the quality of such methods, we then consider what is nowadays referred to as "frequentist Bayesian" or "pragmatic Bayesian" approach which allows to define some notions of optimality for prior choices. We conclude this subsection by presenting some major results obtained in this theory.

Finally, we conclude this overview with two models which illustrate the notions of this overview and which we will study in the following chapter. The first is the inverse Gaussian Sequence Space Model (iGSSM) in section 1.5 and the second is the circular probability density deconvolution section 1.6.

1.1 Inverse problems

We introduce here some fundamentals of inverse problem theory. This section builds upon results which can be found, for example, in Engle et al. (1996).

Consider the situation when one wishes to estimate an object, say f belonging to a space Ξ. The object f will be referred to as "parameter of interest" and the space Ξ as "parameter space". We assume that this parameter has some influence on a system which we are able to observe. Hence, recording observation of this system allows us to learn about this parameter. These observations will be referred to as "data" and denoted by Y. Our ability to learn in such a way is central as it underpins our ability to understand the behaviour of a system, to predict it and to influence it. This is a wide family of problems and we shall give more precision about the specific subfamily we consider.

We will give particular interest to inverse problems, a family of models where one wants to infer on f but the data we observe comes from a system led by a different parameter g which can be written $g := T(f)$ where T is an mapping from Ξ to itself.

These models gathered interest for a long time due to their numerous applications, theoretical physics, astrophysics, medical imaging, econometrics, or acoustics are just a few of the countless examples of such applications. Many of those models have the particularity to be ill-posed in the sense of Hadamard (1902). That is to say, if we build an estimator $\widehat{g}$ of $g = T(f)$ from the data Y and try to apply the inverse T^{-1} of T to this estimator in order to estimate f, one of the following problems might arise:

- non existence (the equation $T(x) = \widehat{g}$ does not have a solution);
- non unicity (the equation $T(x) = \widehat{g}$ has multiple solutions);
- non stability (the solutions to the equations $T(x) = \widehat{g}$ does not depend continuously on $\widehat{g}$).

Though Hadamard thought that inverse problems do not arise in practical situations and that problems of our realm only are of the well-posed kind. Evolution of science proved him wrong and ill-posed problems now have many applications. The specific challenges they represent has since gathered ever increasing interest. We will use two examples throughout this thesis, respectively introduced in section 1.5 and section 1.6.

From now on, we will assume that Ξ is an infinite dimensional vector space on $\mathbb{K}$ (standing for either $\mathbb{R}$ or $\mathbb{C}$), equipped with a norm $\|\cdot\|_\Xi$ which is derived from an inner product $\langle\cdot|\cdot\rangle_\Xi$ and Ξ is hence an infinite dimensional Hilbert space. We denote by $\mathcal{L}(\Xi)$ the set of bounded endomorphisms on Ξ, that is to say linear operators S from Ξ onto itself such that there exists M in $\mathbb{R}_+$ verifying, for any x in Ξ, the following inequality $\|S(x)\|_\Xi \leqslant M\|x\|_\Xi$. In addition, we denote, for any S in $\mathcal{L}(\Xi)$, $\mathcal{D}(S)$ its definition domain, $\mathcal{R}(S)$ its range, and $\mathcal{N}(S)$ its kernel. Assume, from now on, that T is an element of $\mathcal{L}(\Xi)$.

In this case, the following property gives us sufficient and necessary conditions under which the two first forms of ill-posedness do not happen.

Proposition.
For any S in $\mathcal{L}(\Xi)$, and any element x of Ξ, there exists an unique solution to the equation $S(y) = \widehat{S(x)}$ for any estimate $\widehat{S(x)}$ of $S(x)$ in Ξ if and only if

(existence): $\widehat{S(x)}$ *belongs to the range* $\mathcal{R}(S)$ *of* S*;*

(uniqueness): *the operator S is injective, i.e.* $\mathcal{N}(S) = \{0\}$. □

In the case where the existence condition is not fulfilled, one would look for an approximate solution $\widetilde{f}$ minimising an objective function which could be the distance with respect to $\|\cdot\|_\Xi$, that is to say, if it exists, $\widetilde{f} \in \arg\min_{x\in\mathcal{D}(T)} \|T(x) - \widehat{g}\|_\Xi$. If the uniqueness condition is not fulfilled then we can look for the solution with minimal norm, once again, assuming that it exists.
We will see that the orthogonal projection operators, with respect to $\langle\cdot|\cdot\rangle_\Xi$, plays an important role. Indeed, one can show how the last property relates to the orthogonal projection onto the closure of the range of T, $\overline{\mathcal{R}}(T)$. First introduce the following notations.

Definition 1 For any S in $\mathcal{L}(\Xi)$, denote by $S^\star$ its adjoint operator with respect to $\langle\cdot|\cdot\rangle_\Xi$, that is to say the unique operator such that for any x and y in Ξ we have $\langle S(x)|y\rangle_\Xi = \langle x|S^\star(y)\rangle_\Xi$. For any subspace $\mathbb{U}$ of Ξ, denote by $\Pi_\mathbb{U}$ the orthogonal projection onto $\mathbb{U}$ with respect to $\langle\cdot|\cdot\rangle_\Xi$. □

We can now formulate the following property linking the distance minimising criteria with the orthogonal projection onto the closure of the range of T.

Proposition.
For any S in $\mathcal{L}(\Xi)$; any element x of Ξ; any estimate $\widehat{S(x)}$ of $S(x)$ in Ξ; and any estimate $\widetilde{x}$ of x which lies within $\mathcal{D}(S)$, the following assertions are equivalent:

i (distance to the target minimisation) : $\widetilde{x}$ *minimises the function* $y \mapsto \|\widehat{S(x)} - S(y)\|_\Xi$*;*

ii : $\Pi_{\overline{R}(S)}(\widehat{S(x)}) = S(\widetilde{x})$*;*

iii (normal equation) : $S^\star(\widehat{S(x)}) = S^\star(S(\widetilde{x}))$. □

Given those considerations, it is naturally that one defines the generalised inverse (also called pseudo inverse or Moore-Penrose inverse).

Definition 2 For any linear subspace $\mathbb{U}$ of Ξ, denote $\mathbb{U}^\perp$ its orthogonal complement with respect to $\langle\cdot|\cdot\rangle_\Xi$ that is $\mathbb{U}^\perp := \{x \in \Xi : \forall u \in \mathbb{U}, \langle x|u\rangle_\Xi = 0\}$. Moreover, denote $\oplus$ the direct sum binary operator. Then, for any linear operator S, define its generalised inverse S^+ as the unique linear extension of $S^{-1} : \mathcal{R}(S) \to \mathcal{N}(S)^\perp$ to the domain $\mathcal{D}(S^+) := \mathcal{R}(S) \oplus \mathcal{R}(S)^\perp$ with $\mathcal{N}(S^+) = \mathcal{R}(S)^\perp$ satisfying for any x in $\mathcal{D}(S^+)$ the equality $S^+(x) := S^{-1}(\Pi_{\overline{\mathcal{R}}(S)}(x))$. □

One should note that the generalised inverse has the following important properties.

Remark 1.1.1 *For any S in $\mathcal{L}(\Xi)$, the following equalities stand: $SS^+S = S$, $S^+SS^+ = S^+$, $S^+S = \Pi_{\mathcal{N}(S)^\perp}$ and for any x in $\mathcal{D}(S^+)$, $SS^+(x) = \Pi_{\overline{\mathcal{R}}(S)}(x)$. In addition, one should notice that if S is injective, so is $S^\star S$ and as a consequence, $S^\star S : \Xi \to \mathcal{R}(S^\star S)$ is* *invertible which implies that for any x in $\mathcal{R}(S) \oplus \mathcal{R}(S)^{\perp}$ we have that $(S^{\star}S)^{+}S^{\star}x$ is the unique solution of iii (normal equation) which implies that $S^{-1}(\Pi_{\overline{\mathcal{R}(S)}}x) = \{S^{+}x\} = \{(S^{\star}S)^{+}S^{\star}x\}$. Moreover, if S is invertible, S^{+} and S^{-1} coincide.* □

We hence see that the Moore-Penrose inverse offers a solution to the two first sources of ill-posedness.

Proposition.
For any linear operator S from Ξ onto itself and x in $\mathcal{D}(S^{+})$, $S^{+}(x)$ is an element of $S^{-1}(\Pi_{\overline{\mathcal{R}(S)}}x)$ and, hence fulfils i (distance to the target minimisation). Moreover, $S^{+}(x)$ is the unique element fulfilling this condition with minimal $\|\cdot\|_{\Xi}$-norm, that is $\|S^{+}x\|_{\Xi} = \inf\{\|h\|_{\Xi} : h \in S^{-1}(\Pi_{\overline{\mathcal{R}(S)}}x)\}$. □

We will work under a set of assumptions where the two first kinds of ill-posedness do not happen. However, we give more attention to the third source of ill-posedness. The next property gives a general condition under which it occurs.

Proposition.
Let Ξ be infinite dimensional and S be an injective compact linear operator from Ξ onto itself. Then $\inf_{h\in\Xi}\{\|S(h)\|_{\Xi} : \|h\|_{\Xi} = 1\} = 0$ which implies that S^{-1} (and hence S^{+}) are not continuous. □

This discontinuity property highlights the need to define a so called regularised version of the Moore-Penrose inverse. Indeed, it implies that there exists ε in $\mathbb{R}_{+}^{\star}$ such that for any δ in $\mathbb{R}_{+}^{\star}$, there exists a couple (x, y) of elements of Ξ with $\|x - y\|_{\Xi} \leqslant \delta$, such that $\|S^{+}(x) - S^{+}(y)\|_{\Xi} \geqslant \varepsilon$. Taking $x = g$ and $(y_n)_{n\in\mathbb{N}} = (\widehat{g}_n)_{n\in\mathbb{N}}$ a sequence of estimators, it means that even if $(\widehat{g}_n)_{n\in\mathbb{N}}$ is a consistent sequence of estimations for g, $S^{+}(\widehat{g})$ would still not be a consistent estimator of f.

We will see later in this overview that depending on the approach one uses, the strategy to overcome this difficulty will not be the same. Namely, in the frequentist paradigm, one introduces the notion of regularisation in order to define a continuous approximation of T^{+} whereas in the Bayesian paradigm, this regularisation occurs naturally in this derivation of the posterior distribution.
To make this clearer, we will first introduce the shape that our data will take.

1.2 Data types

In the previous section we gave details about the nature of the object we want to estimate, f; the object which we gather information about, g; as well as the operator which links them, T, that is to say, $g = T(f)$. However, we only loosely mentioned the data Y which we gather and the estimate $\widehat{g}$ of g it allows us to construct. Throughout this thesis our data will be regarded as Ξ-indexed stochastic processes for which we give the definition hereafter.

Definition 3 Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let be $\mathcal{B}$, the Borel sigma-algebra on $\mathbb{K}$. Consider a family of $\mathbb{K}$-valued random variables indexed by Ξ, say $\{X(x), x \in \Xi\}$, that is to say, for any x in Ξ, $X(x)$ is a measurable mapping from $(\Omega, \mathcal{A})$ to $(\mathbb{K}, \mathcal{B})$. Then we call Ξ-indexed stochastic process the mapping $X : (\Omega, \mathcal{A}) \to (\mathbb{K}^{\Xi}, \mathcal{B}^{\otimes \Xi}), \quad \omega \mapsto (X(x)(\omega))_{x \in \Xi}$. □

Hence, to ease the study of stochastic processes we introduce the following notations for $(\mathbb{K}, \mathcal{B})$-valued random variables.

Definition 4 For any z in $\mathbb{K}$, let us denote $\overline{z}$ its complex conjugate, hence for $\mathbb{K} = \mathbb{R}$ we have $z = \overline{z}$. For any random variable $X : (\Omega, \mathcal{A}) \to (\mathbb{K}, \mathcal{B})$, let $\mathbb{P}_X$ be the measure on $(\mathbb{K}, \mathcal{B})$ given, for any B in $\mathcal{B}$, by $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$. The set $\mathbb{L}^2(\Omega)$ of square integrable random variables can be defined by $\{X : (\Omega, \mathcal{A}) \to (\mathbb{K}, \mathcal{B}), \int_{\mathbb{K}} |t|^2 \, d\mathbb{P}_X(t) < \infty\}$. On this set, the expectation, variance, and covariance operators $\mathbb{E}$, $\mathbb{V}$, and $\mathbb{C}\mathrm{ov}$ can properly be defined and are given by

$$\begin{aligned} \mathbb{E} \quad &: \mathbb{L}^2(\Omega) \to \mathbb{K}; & \mathbb{V} &: \mathbb{L}^2(\Omega) \to \mathbb{R}_+; \\ &\quad X \mapsto \int_{\mathbb{K}} t \, d\mathbb{P}_X(t) & &\quad X \mapsto \mathbb{E}[|X|^2] - |\, \mathbb{E}[X]|^2 \\ \mathbb{C}\mathrm{ov} &: (\mathbb{L}^2(\Omega))^2 \to \mathbb{R}; & & \\ &\quad (X, Z) \mapsto \mathbb{E}[X\overline{Z}] - \mathbb{E}[X]\,\mathbb{E}[\overline{Z}] & & \end{aligned}$$

where the integrals are in the sense of Lebesgue.

Given a family of probability distributions indexed by Ξ, denoted $(\mathbb{P}_x)_{x \in \Xi}$, a function S defined on $\mathbb{K}$, and x in Ξ, we denote $\mathbb{E}_x[S(X)] = \int_{\mathbb{K}} f(s) \, d\mathbb{P}_x(s)$ and $\mathbb{V}_x[S(X)] = \mathbb{E}_x[|S(X)|^2] - |\, \mathbb{E}_x[S(X)]|^2$, the expected value and variance of the random variable $S(X)$ if X admits $\mathbb{P}_x$ as a distribution. □

We can now formulate properly the notion of mean process and covariance process for a stochastic process.

Definition 5 Given a Ξ-indexed stochastic process, say $X = (X(x))_{x \in \Xi}$, such that, for any x in Ξ, $X(x)$ is an element of $\mathbb{L}^2(\Omega)$. Then, the mean function of X is the mapping from Ξ to $\mathbb{K}$, such that given by $\mathbb{E}[X] = (\Xi \to \mathbb{K}, \quad x \mapsto \mathbb{E}[X(x)])$, and the covariance operator is given by by $\mathbb{C}\mathrm{ov}[X] = (\Xi^2 \to \mathbb{K}), \quad (x, y) \mapsto \mathbb{C}\mathrm{ov}(X(x), X(y))$. □

Keeping those definitions in mind, we will consider two configurations for our data. One when T is known and we have at hand a sample allowing to estimate $T(f)$ and the other when T is unknown and we have two samples at hand, one to estimate T and the other to estimate $T(f)$.
In any case, considering the models which we will use as illustrations as well as the technics we will use, introducing the following hypotheses and notations will be of great use.

Definition 6 Let $\mathbb{F}$ be either $\mathbb{N}$ or $\mathbb{Z}$ which will we will refer to as frequency domain. Then, let $\mathcal{U} := (e_s)_{s \in \mathbb{F}}$ be an orthonormal system of Ξ indexed by $\mathbb{F}$, that is to say a family of elements of Ξ such that for any two elements of $\mathbb{F}$, s_1 and s_2, we have $\langle e_{s_1} | e_{s_2} \rangle_{\Xi} = \mathbb{1}_{\{s_1 = s_2\}}$ where for any assertion A, $\mathbb{1}_A$ stands for the function of A which is equal to 1 if A is true and 0 otherwise. In addition, denote by μ the counting measure on $\mathbb{F}$. We denote $\mathbb{U}$ the linear space spanned by $\mathcal{U}$. □

REMARK 1.2.1 *One could consider the case where $\mathbb{F}$ is $\mathbb{R}$ and use the Lebesgue measure as μ, however, such considerations are beyond the scope of this thesis.*

In the examples considered in this thesis the following hypothesis holds true.

ASSUMPTION 1 The parameter of interest f lies in $\mathbb{U}$. □

Once we found such an infinite dimensional linear subspace of Ξ for which we have an orthonormal basis we can consider the generalised Fourier transform and base our inference on the Fourier space. It is in this perspective that we give the following definitions.

DEFINITION 7 Denote Θ the space of mappings from $\mathbb{F}$ onto $\mathbb{K}$. Equipped with the usual addition $+ : \Theta^2 \to \Theta, \quad ([x],[y]) \mapsto ([x]+[y] : s \mapsto [x](s)+[y](s))$ and external product $\cdot : \mathbb{K} \times \Theta \to \Theta, \quad (a,[y]) \mapsto (a \cdot [y] : s \mapsto a \cdot [y](s))$ it is a linear vector space. In addition, defining the conjugate of $[x]$, $\overline{[x]}$ such that for any s in $\mathbb{F}$ we have $\overline{[x]}(s) = \overline{[x](s)}$, we may define the following inner product: $\langle \cdot | \cdot \rangle_\Theta : ([x],[y]) \mapsto \langle [x] | [y] \rangle_\Theta = \sum_{s \in \mathbb{F}} [x](s) \cdot \overline{[y]}(s)$. Hence $(\Theta, \langle \cdot | \cdot \rangle_\Theta)$ is an Hilbert space. □

With those objects at hand, we define the generalised Fourier transform on Ξ.

DEFINITION 8 Define the generalised Fourier transform linear operator $\mathcal{F}$ by $\mathcal{F} : \mathbb{U} \to \Theta, \quad x \mapsto \mathcal{F}(x) := (s \mapsto ([x](s) := \langle x | e_s \rangle_\Xi))$. We see that $\mathcal{F}$ is a unitary linear mapping between Hilbert spaces and we should highlight that its conjugate (which is hence also its inverse) is given by $\mathcal{F}^\star : \Theta \to \Xi; \quad [x] \mapsto \sum_{s \in \mathbb{F}} [x](s) e_s$. □

With those definitions at hand we formulate the following hypothesis about T.

ASSUMPTION 2 We assume that, for any s in $\mathbb{F}$, there exist an element of $\mathbb{K} \setminus \{0\}$, say $\lambda(s)$ such that $\langle T(e_s) | e_s \rangle_\Xi = \lambda(s)$. In other words, $(e_s)_{s \in \mathbb{F}}$ diagonalises T and we have, for any x in $\mathbb{U}$ that $T(x) = \int_{\mathbb{F}} \lambda(s)[x](s) e_s \, \mathrm{d}\mu(s)$. □

Following naturally from the definitions and hypotheses we just introduced, we will use the following notations.

DEFINITION 9 Let θ°, λ, and ϕ be the elements of Θ such that for any s in $\mathbb{F}$ we have

$$\theta^\circ(s) := \mathcal{F}(f)(s); \quad \lambda(s) := \langle T(e_s) | e_s \rangle_\Xi; \quad \phi(s) := \mathcal{F}(T(f))(s) = \mathcal{F}(g)(s).$$

In addition let h be the element of Ξ such that $h := \mathcal{F}^\star(\lambda)$. □

Notice that, as $g = T(f)$, for any s in $\mathbb{F}$, we have $\phi(s) = \theta^\circ(s)\lambda(s)$.
Considering a Ξ-indexed stochastic process $(Y(x))_{x \in \Xi}$, and in particular its sub-process $(Y(e_s))_{s \in \mathbb{F}}$, which is hence a $\mathbb{F}$-indexed stochastic process, we can define a distribution on Ξ considering the random variable $X : (\Omega, \mathcal{A}) \to (\Xi, \mathcal{B}), \quad \omega \mapsto \mathcal{F}^\star((Y(e_s)(\omega))_{s \in \mathbb{F}})$. Reciprocally, considering a Ξ-valued random variable X, one can define a Ξ-indexed stochastic process $(Y(x))_{x \in \Xi}$ where, for any x in Ξ, $Y(x)$ is the random variable defined by $Y(x) = \langle X | x \rangle_\Xi$ and in particular one can define the $\mathbb{F}$-indexed process $(Y(s))_{s \in \mathbb{F}}$ where $Y(s) = \mathcal{F}(X)(s)$. One can then notice that for any x in Ξ, $Y(x) = \langle X | x \rangle_\Xi = \sum_{s \in \mathbb{F}} \overline{[x]}(s) \langle X | e_s \rangle_\Xi = \sum_{s \in \mathbb{F}} \overline{[x]}(s) Y(s)$.
We can now give a more precise shape for our observations which will come in two flavours described in the two following subsections.

1.2.1 Ill-posed inverse problem with known operator

In this first case, given a sequence of Ξ-valued random variables indexed by $\mathbb{Z}$, say $(Y_p)_{p\in\mathbb{Z}}$ and an integer integer n, our observation Y^n is assumed to be $(Y_p)_{p\in[\![1,n]\!]}$, where, for any two a and b in $\mathbb{Z}$, $[\![a,b]\!]$ stands for $[a,b]\cap\mathbb{Z}$. That is, there exists a σ-algebra $\mathcal{B}$ on Ξ such that, for any p in $\mathbb{Z}$, Y_p is a measurable mapping from $(\Omega,\mathcal{A})$ to $(\Xi,\mathcal{B})$. Then, $\Omega\to\mathbb{K}^{\Xi\times\mathbb{Z}}, \quad \omega\mapsto(\langle Y_p|x\rangle_\Xi)_{p\in\mathbb{Z},x\in\Xi}$ is a stochastic process on $\Xi\times\mathbb{Z}$.
The inference on f is then based on the following assumption.

Definition 10 Consider $(Y_p)_{p\in\mathbb{Z}}$, a Ξ-valued stochastic process. It is called strictly stationary if, for any r in $\mathbb{Z}$, q in $\mathbb{N}$, and $(p_i)_{i\in[\![1,q]\!]}$ in $\mathbb{Z}^q$, the vectors of random variables $(Y_{p_i})_{i\in[\![1,q]\!]}$ and $(Y_{p_i+r})_{i\in[\![1,q]\!]}$ are identically distributed. In such a process, the marginals are obviously identically distributed and we denote $\mathbb{P}_Y := \mathbb{P}\circ Y_0^{-1}$ the distribution of the marginals. □

Assumption 3 Assume that the operator T is known and $(Y_p)_{p\in\mathbb{Z}}$ is strictly stationary. In addition, assume that the distribution of Y belongs to a family indexed by Ξ, denoted $(\mathbb{P}_x)_{x\in\Xi}$ and that $Y\sim\mathbb{P}_g$ where, for any $(\Xi,\mathcal{B})$-valued random variable X and measure $\mathbb{Q}$ on $(\Xi,\mathcal{B})$, $X\sim\mathbb{Q}$ means that for any B in $\mathcal{B}$, $\mathbb{P}\circ X^{-1}(B)=\mathbb{Q}(B)$. We assume that for any z in Ξ and random variable Z such that $Z\sim\mathbb{P}_z$, we have for any y in Ξ, $\mathbb{E}[|\langle Z|y\rangle_\Xi|^2]<\infty$; and, in particular, $\mathbb{E}[\langle Z|y\rangle_\Xi]=\langle z|y\rangle_\Xi$. □

A direct consequence of this hypothesis is that for any s in $\mathbb{F}$, we have $\mathbb{E}[\langle Y|e_s\rangle_\Xi] = \langle g|e_s\rangle_\Xi = \phi(s)$. Due to the invertible nature of $\mathcal{F}$, we will indifferently denote $(\mathbb{P}_x)_{x\in\mathbb{U}}$ and $(\mathbb{P}_{[x]})_{[x]\in\Theta}$ with the identification, for any x in $\mathbb{U}$, $\mathbb{P}_x=\mathbb{P}_{\mathcal{F}(x)}$. In particular, we have $\mathbb{P}_g=\mathbb{P}_\phi$, $\mathbb{P}_f=\mathbb{P}_{\theta^\circ}$, and $\mathbb{P}_h=\mathbb{P}_\lambda$. Generally, we will denote Y, X, and ε, random variables with respective distributions $\mathbb{P}_g$, $\mathbb{P}_f$, and $\mathbb{P}_h$ or equivalently $\mathbb{P}_\phi$, $\mathbb{P}_{\theta^\circ}$, and $\mathbb{P}_\lambda$.

1.2.2 Ill-posed inverse problem with unknown operator

Similarly to the previous case, we still observe replications $(Y_p)_{p\in[\![1,n]\!]}$ of a stochastic process Y, however, we also observe replications $(\varepsilon_q)_{q\in[\![1,n_\lambda]\!]}$ of a second Ξ-valued stochastic process ε. This second set of observation is used to estimate T which is not considered as known anymore.

Assumption 4 Assume that $(Y_p)_{p\in\mathbb{Z}}$ and $(\varepsilon_p)_{p\in\mathbb{Z}}$ are strictly stationary. In addition, assume that the distributions of the marginals in $(Y_p)_{p\in\mathbb{Z}}$ and $(\varepsilon_p)_{p\in\mathbb{Z}}$ belong to a family indexed by Ξ, denoted $(\mathbb{P}_x)_{x\in\Xi}$ such that, for any p in $\mathbb{Z}$, $Y_p\sim\mathbb{P}_g$ and $\varepsilon_p\sim\mathbb{P}_h$. □

1.2.3 Independent data

In the two previous subsections, we have described the mean function of the two processes we observe. However, we haven't discussed the covariance operator, except by assuming that the diagonal is finite.
We will consider two assumptions for the dependence structure. The first is independence.

Assumption 5 We assume that, for any m in $\mathbb{N}$, and vector $(p_q)_{q\in[\![1,m]\!]}$ in $\mathbb{Z}^m$, Y_{p_q} is an independent vector. That is to say, for any $(B_q)_{q\in[\![1,m]\!]}$ in $\mathcal{B}^m$, $\mathbb{P}(\cap_{q=1}^m Y_{p_q}^{-1}(B_q)) = \prod_{q=1}^m \mathbb{P}(Y_{p_q}^{-1}(B_q))$. □

Among other things, this implies that for any p and q, with $p \neq q$ in $\mathbb{Z}$ and x and y in Ξ, $\mathbb{E}[\langle Y_p|x\rangle_\Xi \cdot \langle Y_q|y\rangle_\Xi] = \mathbb{E}[\langle Y_p|x\rangle_\Xi] \cdot \mathbb{E}[\langle Y_q|y\rangle_\Xi]$ which also implies $\mathbb{V}[\langle Y_p|x\rangle_\Xi + \langle Y_q|y\rangle_\Xi] = \mathbb{V}[\langle Y_p|x\rangle_\Xi] + \mathbb{V}[\langle Y_q|y\rangle_\Xi]$.
In this case, $(Y_p)_{p\in\mathbb{Z}}$ is a sequence of independent identically distributed (i.i.d.) random variables.

1.2.4 Absolutely regular process

Even though the independence assumption is widely spread, it is also limiting as, in practice, dependent data arise often. Hence, the inference based on dependent data gathered a lot of interest in the past and it appears clearly that one should limit the degree of dependence which is permitted in order to obtain theoretical results as well as technics which perform properly.
We hence first introduce the notion of beta mixing coefficients which allows a quantification of dependence.

DEFINITION 11 β-MIXING COEFFICIENTS
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{U}$ and $\mathcal{V}$ be two sub σ-algebras of $\mathcal{A}$. Then, we define the β-mixing coefficient of $\mathcal{U}$ and $\mathcal{V}$:

$$\beta(\mathcal{U}, \mathcal{V}) := \frac{1}{2} \sup_{(U_j)_{j\in I}(V_j)_{j\in J}} \left\{ \sum_{j\in J} \sum_{k\in I} |\,\mathbb{P}(U_j)\,\mathbb{P}(V_k) - \mathbb{P}(U_j \cap V_k)| \right\}$$

where the sup is taken over all possible finite partition of Ω which are respectively $\mathcal{U}$ and $\mathcal{V}$ measurable.
In addition for two random variables Z_1 and Z_2 we note $\sigma(Z_1)$ and $\sigma(Z_2)$ the σ-algebra they generate and $\beta\,(Z_1, Z_2) = \beta\,(\sigma(Z_1), \sigma(Z_2))$. □

With this definition at hand, one can define an absolutely regular process which is a stochastic process for which the beta mixing coefficients fade for increasingly distant observations.

DEFINITION 12 ABSOLUTELY REGULAR PROCESS
Consider a stochastic process $(Z_p)_{p\in\mathbb{Z}}$. Denote, for any p in $\mathbb{N}$, by $\mathcal{F}_p^- := \sigma\left((Z_q)_{q\leqslant p}\right)$ and $\mathcal{F}_p^+ := \sigma\left((Z_q)_{q\geqslant p}\right)$. The stochastic process $(Z_p)_{p\in\mathbb{Z}}$ is said to be absolutely regular if

$$\lim_{p\to\infty} \beta(\mathcal{F}_0^-, \mathcal{F}_p^+) = 0.$$

□

An interesting result using this definition, which can be found in this form in Asin and Johannes (2016) and is adapted from THEOREM 2.1 in Viennet (1997), links the β-mixing coefficients of a stochastic process and its variance.

LEMMA 1.2.1.
Let $(Z_p)_{p\in\mathbb{Z}}$ be a $\mathbb{R}$-valued, strictly stationary, stochastic process. There exists a sequence $(b_p)_{p\in\mathbb{N}}$ of measurable functions from $\mathbb{R}$ to $[0,1]$ with, for any p in $\mathbb{N}$, $\mathbb{E}[b_p(Z_0)] = \beta(Z_0, Z_p)$ such that, for any measurable function x such that $\mathbb{E}[|x(Z_0)|^2] < \infty$ and any integer n, we have $\mathbb{V}[\sum_{p=1}^n x(Z_p)] \leqslant n\,\mathbb{E}[|x(Z_0)|^2(1 + 4\sum_{p=1}^{n-1} b_p(Z_0))]$. □

Notice that, in this lemma, if x is a bounded function, say $\|x\|_\infty := \sup_{t\in\mathbb{R}} |x(t)|^2 \leqslant 1$, then we obtain $\mathbb{V}[\sum_{p=1}^n x(Z_p)] \leqslant n\,\mathbb{E}[1 + 4\sum_{p=1}^{n-1} b_p(Z_0)]$. In our context, x will generally be an element of a function basis, such as the complex exponential trigonometric basis, which is hence indeed bounded. Notice that the bound we obtain here depends on the sequence of β-coefficients which are generally unknown and their estimation is a challenging task. Hence, while this bound is sufficient to obtain convergence rates for non adaptive estimators (as it will be formulated explicitly further), it is in general necessary to give a stronger hypothesis on the observation process in order to obtain properties for sophisticated adaptive methods. In this optic, let's introduce the following space of functions.

Definition 13 Given $q \geqslant 2$, a non negative sequence $\omega = (\omega_p)_{p\in\mathbb{N}}$ and a probability measure $\mathbb{P}$, let $\mathcal{L}(q, \omega, \mathbb{P})$ be the set of functions $b : \mathbb{R} \to \overline{\mathbb{R}_+}$ such that there exists a sequence $(b_p)_{p\in\mathbb{N}}$ of measurable functions $b_p : \mathbb{R} \to [0, 1]$ with $b_0 : x \mapsto 1$ and, for any random variable Z such that $Z \sim \mathbb{P}$, we have $\mathbb{E}[b_p(Z)] \leqslant \omega_p$ satisfying $b = \sum_{p=0}^\infty (p + 1)^{q-2} b_p$. □

One can easily see that a sufficient condition for elements of $\mathcal{L}(q, \omega, \mathbb{P})$ to be non-negative $\mathbb{P}$-integrable functions is $\sum_{p=0}^\infty (p-1)^{q-2}\omega_p < \infty$. Combined with Lemma 1.2.1, we obtain the following lemma.

Lemma 1.2.2.

Let $(Z_p)_{p\in\mathbb{Z}}$ be a $\mathbb{R}$-valued, strictly stationary, stochastic process with common marginal distribution $\mathbb{P}_Y$. Denote $(\omega_p)_{p\in\mathbb{N}}$ the sequence of β-mixing coefficients. There exists a function b in $\mathcal{L}(2, \omega, \mathbb{P}_Y)$ such that, for any measurable function x such that $\mathbb{E}[|x(Y_0)|^2] < \infty$ and any integer n, we have $\mathbb{V}[\sum_{p=1}^n x(Z_p)] \leqslant 4n\,\mathbb{E}[|x(Z_0)|^2 b(Z_0))]$.
Alternatively, assuming, for r and q exponents as in Hölder's inequality, that β_p tends to 0 as p tends to ∞ with $\omega_0 = 1$ and that, for some r, $\sum_{p\in\mathbb{N}}(p + 1)^{r-1}\beta_p < \infty$ then, we have

$$\mathbb{E}\left[|x(Z_0)|^2\, b(Z_0)\right] \leqslant \mathbb{E}\left[|x(Z_0)|^{2q}\right]^{1/q} \left(r \sum\nolimits_{p\in\mathbb{N}} (p + 1)^{r-1}\beta_p\right)^{1/r}.$$

□

Notice, once again, that with $\|x\|_\infty = 1$ we have $\mathbb{V}[\sum_{p=1}^n x(Z_p)] \leqslant 4n \sum_{p\in\mathbb{N}} \beta(Z_0, Z_p)$ which implies, jointly with the assumption "$\sum_{p\in\mathbb{N}} \beta(Z_0, Z_p) < \infty$", there exists a constant $\mathcal{C}$ such that $\mathbb{V}[\sum_{p=1}^n x(Z_p)] \leqslant \mathcal{C}n$. Notice, though, that this bound would depend on a constant related to the β-mixing coefficients. It hence allows to show that for sequences $\beta(Z_0, Z_p)$ decreasing sufficiently fast, the oracle or minimax risk is the same as for independent sequences, however, it remains unsuitable for the study of adaptive methods. We hence present a third inequality which relies on the following assumption, regularly used in the study of such processes, for example in Asin and Johannes (2016, 2017); Bosq (2012).

Assumption 6 Considering a $[0, 1]$-valued stochastic process $(Z_p)_{p\in\mathbb{Z}}$, assume that for any p, the joint distribution $\mathbb{P}_{Z_0,Z_p}$ of Z_0 and Z_p admits a density denoted x_{Z_0,Z_p} which is square integrable. Denote the L^2-norm for functions of two variables by $\left\|x_{Z_0,Z_p}\right\|_{L^{2,2}}^2 := \int\int_{\mathbb{K}^2} |x_{Z_0,Z_p}(t_0, t_p)|^2\, \mathrm{d}t_0\, \mathrm{d}t_p$ and for any t_0 and t_p in $[0, 1]$ set $(x \otimes x)(t_0, t_p) = x(t_0) \cdot x(t_p)$. Then, we assume $\gamma_x := \sup_{p\geqslant 1} \|x_{Z_0^n,Z_p^n} - x \otimes x\|_{L^{2,2}} < \infty$. □

Lemma 1.2.3.
Let the process $(Z_p)_{p\in\mathbb{N}}$ be a strictly stationary process with associated sequence of mixing coefficients $(\beta(Z_0,Z_p))_{p\in\mathbb{N}}$ verifying Assumption 6 and a sequence of functions e_s from $[0,1]$ to $\mathbb{K}$ such that $\|e_s\|_{L^\infty}=1$. Then, for any $n\geqslant 1$; m and l in $\mathbb{N}$ with $m\leqslant l$ and $K\in[\![0,n-1]\!]$, it holds

$$\begin{aligned}\sum_{m\leqslant|s|\leqslant l}\mathbb{V}\left[\sum_{p=1}^{n}e_s(Z_p)\right]&\\ \leqslant n2(l-m+1)&\left\{1+2\left[\gamma_x K(l-m+1)^{-1/2}+2\sum_{p=K+1}^{n-1}\beta(Z_0,Z_p)\right]\right\}.\end{aligned}$$

Moreover, as $\sum_{p\in\mathbb{N}}\beta(Z_0,Z_p)$ is finite, we have $\lim_{K\to\infty}\sum_{p=K+1}^{\infty}\beta(Z_0,Z_p)=0$, so we can find K° in $\mathbb{N}$ such that for any K greater than K°, $\sum_{p=K+1}^{\infty}\beta(Z_0,Z_p)\leqslant\frac{1}{4}$. We can take $K=\frac{\sqrt{l-m+1}}{4\gamma_x}$ and assuming that this choice is greater than K°, we have

$$\sum_{m\leqslant|s|\leqslant l}\mathbb{V}\left[\sum_{p=1}^{n}e_s(Z_p)\right]\leqslant 4n(l-m+1).$$

□

Contrarily to the previous lemmata, this one exhibits an upper bound for the variance which does not involve the sum of mixing coefficients which allows to design a data driven estimator which does not requires knowledge of them. Finally, to use this last lemma properly, we will need one last result, which can be found in Viennet (1997).

Lemma 1.2.4.
Assume that the universe is rich enough in the sense that there exist a sequence of random variables with uniform distribution on $[0,1]$ which is independent of $(Z_p)_{p\in\mathbb{Z}}$.
Then, there exist a sequence $(Z_p^\perp)_{p\in\mathbb{Z}}$ satisfying the following properties. For any positive integer w and for any strictly positive integer q, define the sets $(I^e_{q,p})_{p\in[\![1,w]\!]}:=[\![2(q-1)w+1,(2q-1)w]\!]$ and $\left(I^o_{q,p}\right)_{p\in[\![1,w]\!]}:=[\![(2q-1)w+1,2qw]\!]$.
Define for any q in $\mathbb{Z}$ the vectors of random variables $E_q:=(Z^n_{I^e_{q,p}})_{p\in[\![1,w]\!]}$; $O_q:=(Z^n_{I^o_{q,p}})_{p\in[\![1,w]\!]}$; and their counterparts $E_q^\perp:=(Z^{n,\perp}_{I^e_{q,p}})_{p\in[\![1,w]\!]}$ and $O_q^\perp:=(Z^{n,\perp}_{I^o_{q,p}})_{p\in[\![1,w]\!]}$.
Then, $\left(Z_p^\perp\right)_{p\in\mathbb{N}}$ satisfies:

- *for any integer q, $E_q^\perp$, E_q, $O_q^\perp$, and O_q are identically distributed;*
- *for any integer q, $\mathbb{P}^n_{\theta^\circ}\left(E_q\neq E_q^\perp\right)\leqslant\beta_w$ and $\mathbb{P}^n_{\theta^\circ}\left(O_q\neq O_q^\perp\right)\leqslant\beta_w$;*
- *$\left(E_q^\perp\right)_{q\in\mathbb{Z}}$ are independent and identically distributed and $\left(O_q^\perp\right)_{q\in\mathbb{Z}}$ as well.*

□

Note that, even though this is the only quantification of dependence we will consider in this thesis, many other have been considered and overviews can be found in Bosq (2012); Bradley (2005).

1.3 Frequentist approach

In the two previous sections, we have first introduced inverse problems in a general context and highlighted some difficulties which are inherent to this kind of problem. We then introduced the type of data we will have at hand. Now, we aim to introduce the methods we will use and more generally the paradigm they conform to, what motivates their construction and how to justify satisfaction or dissatisfaction regarding their properties. As explained in the introduction, our methods will be of two kinds, namely frequentist and Bayesian. In this section, we present the frequentist paradigm and the notions of decision theory which allow to quantify the quality of frequentist estimation methods.

1.3.1 Estimation

Remind that, given a family of probability distributions on Ξ and indexed by Ξ itself $(\mathbb{P}_x)_{x\in\Xi}$ we are interested in estimating an object f in Ξ while observing some data Y from $\mathbb{P}_g$ where $g = Tf$ with T a linear operator from Ξ onto itself. Then, the frequentist approach consists in defining an estimator of the parameter of interest using the data where an estimator is an application as defined hereafter.

Definition 14 Given a parameter space $(\Xi, \mathcal{A})$ and an observation space $(\mathbb{Y}, \mathcal{Y})$, an estimator is a measurable application from $(\mathbb{Y}, \mathcal{Y})$ to $(\Xi, \mathcal{A})$. □

Hence, in our particular case, an estimator would be any measurable application from $(\Xi^n, \mathcal{A}^{\otimes n})$ to $(\Xi, \mathcal{A})$. As mentioned earlier, using the generalised Fourier transform, we will go through the space of sequences Θ, equipped with the Borel sigma algebra generated by the l^2-norm, say $\mathcal{B}$. In our context, some naive estimators for relevant objects of the model we consider are the so-called "empirical estimators" or "orthogonal series estimator" (OSE).

Definition 15 Keeping in mind that we observe $Y^n = (Y_p)_{p\in[\![1,n]\!]}$ where $(Y_p)_{p\in\mathbb{Z}}$ is a stationary process such that, for any p in $\mathbb{Z}$, we have Y_p follows $\mathbb{P}_g$, where $(\mathbb{P}_x)_{x\in\Xi}$ is a probability distribution on Ξ. Define, for any s in $\mathbb{F}$

$$\begin{aligned} \phi_n(s) :&(\Xi^n, \mathcal{A}^{\otimes n}) \to (\Theta, \mathcal{B}); & \theta_n(s) :&(\Xi^n, \mathcal{A}^{\otimes n}) \to (\Theta, \mathcal{B}); \\ &Y^n \mapsto n^{-1}\sum_{p=1}^{n} \langle Y_p | e_s\rangle_\Xi & &Y^n \mapsto \phi_n(s)\lambda^{-1}(s) \end{aligned}$$

where λ^{-1} is well defined as we assumed $\lambda(s) \neq 0$ for any s. If it were not the case, one would use the generalised inverse $\lambda^{+}(s) = \lambda(s)^{-1}\mathbb{1}_{\{\lambda(s)\neq 0\}}$.
Note that this definition is suitable under assumption Assumption 3 but not Assumption 4 as it relies on the knowledge of λ to be computed; in this case we would consider

$$\begin{aligned} \theta_{n,n_\lambda}(s) :&(\Xi^{n+n_\lambda}, \mathcal{A}^{\otimes(n+n_\lambda)}) \to (\Theta, \mathcal{B}); \\ &(Y^n, \varepsilon^{n_\lambda}) \mapsto \phi_n(s)\lambda^{+}_{n_\lambda}(s) \end{aligned}$$

where we define, for any s in $\mathbb{F}$ the estimator $\lambda_{n_\lambda}(s) := n_\lambda^{-1}\sum_{p=1}^{n_\lambda}\langle\varepsilon_p|e_s\rangle_\Xi$ of $\lambda(s)$ and $\lambda^{+}_{n_\lambda}(s) := \mathbb{1}_{\{|\lambda_{n_\lambda}(s)|^2 > n_\lambda^{-1}\}}\lambda^{-1}_{n_\lambda}(s)$ which hence does not rely on the knowledge of λ but the

information we have about it through the observation of ε^{n_λ}.
From these estimators one can naturally build their counterparts

$$g_n : Y^n \mapsto \mathcal{F}^{-1}(\phi_n); \quad h_{n_\lambda} : \varepsilon^{n_\lambda} \mapsto \mathcal{F}^{-1}(\lambda_{n_\lambda}); \quad f_{n,n_\lambda}; (Y^n, \varepsilon^{n_\lambda}) \mapsto \mathcal{F}^{-1}(\theta_{n,n_\lambda}).$$

□

However, we have seen in section 1.1 that inverse problems define a class of statistical models which has three major characteristics. We have also seen that two of them (non-existence or non-unicity of the solution) can be addressed thanks to the generalised inverse construction. However, we also pointed out that even once one has addressed those two issues, they can still face the difficulty of instability of the solution.
We will see that the estimators we just defined do not escape this phenomenon.
It is in order to address this issue that one defines the family of operators called regularisations.

Definition 16 Given S in $\mathcal{L}(\Xi)$, a family of elements of $\mathcal{L}(\Xi)$, say $\{S_m^+, m \in \mathbb{R}_+\}$ is called regularisation of S^+ if, for any x in $\mathcal{D}(S^+)$ holds $\lim_{m\to\infty} \|S_m^+ x - S^+ x\|_\Xi = 0$. □

Note that the definition of such a family does not solve the problem by itself. Indeed, define the operator norm such that, for any S in $\mathcal{L}(\Xi)$ we have, $\|S\|_{\mathcal{L}(\Xi)} := \sup\{\|S(x)\|_\Xi, x \in \Xi, \|x\|_\Xi \leqslant 1\}$. Then, if S^+ is not bounded, then, for any regularisation of S^+, we have $\lim_{m\to\infty} \|S_m^+\|_{\mathcal{L}(\Xi)} = \infty$ and hence the limit itself is not an element of $\mathcal{L}(\Xi)$.
However, for any S in $\mathcal{L}(\Xi)$, and x in Ξ, if we have a sequence of estimates, indexed by an integer n, say, $(\widehat{S(x)}_n)_{n\in\mathbb{N}}$ of $S(x)$ such that $\lim_{n\to\infty} \|\widehat{S(x)}_n - S(x)\|_\Xi = 0$, then, there exist a sequence m_n such that $\lim_{n\to\infty} \|S_{m_n}^+(\widehat{S(x)}_n) - S^+(S(x))\|_\Xi = 0$ and hence there exists a consistent estimation procedure.
Hence, we see that the selection of the parameter m, which we will call regularisation parameter, is primordial. Depending on it, the estimation procedure could be consistent or not. In addition, within the choices leading to consistent estimation, one can obtain various convergence rates.
In this thesis, the so called regularisation by dimension reduction plays a central role.
The regularisation consists in projecting our estimate onto the "lower frequencies" from $\mathcal{U}$. To do so, consider the following definition.

Definition 17 Consider an index set $\mathbb{M}$ (here $\mathbb{N}$), and a sequence of measurable subsets of $\mathbb{F}$ indexed by $\mathbb{M}$, say, $(\mathbb{F}_m)_{m\in\mathbb{M}}$. This sequence is called a nested sieve if:

i: for any k and m in $\mathbb{M}$ such that $k \leqslant m$, we have $\mathbb{F}_k \subset \mathbb{F}_m$;

ii: for any m in $\mathbb{M}$, we have $\mu(\mathbb{F}_m) < \infty$;

iii: $\cup_{m\in\mathbb{M}} \mathbb{F}_m = \mathbb{F}$. □

Similarly, for any m in $\mathbb{M}$, we define $\mathbb{U}_{\overline{m}}$ the linear subspace of $\mathbb{U}$ generated by $(e_s)_{s\in\mathbb{F}_m}$.
For any m in $\mathbb{M}$, we will denote the set $\mathbb{F} \setminus \mathbb{F}_m$ by $\mathbb{F}_m^c$. In all the examples in this thesis, $\mathbb{F}$ will be either $\mathbb{N}$ or $\mathbb{Z}$; $\mathbb{M}$ will be $\mathbb{N}$; and for any m in $\mathbb{N}$, $\mathbb{F}_m$ will be $\{s \in \mathbb{F} : |s| \leqslant m\}$.
The following notation will hence be regularly used: for any s_1 and s_2 in $\mathbb{Z}$ with $s_1 \leqslant s_2$ we denote $[\![s_1, s_2]\!]$ the set $[s_1, s_2] \cap \mathbb{Z}$.

By extension, for any m_1 and m_2 in $\mathbb{M}$, we will denote $\mathbb{U}_{\underline{m_1}}$ the linear subspace of $\mathbb{U}$ generated by $(e_s)_{s\in\mathbb{F}^c_{m_1}}$; and $\mathbb{U}_{\underline{m_1},\overline{m_2}}$ the linear subspace of $\mathbb{U}$ generated by $(e_s)_{s\in\mathbb{F}^c_{m_1}\cap\mathbb{F}_{m_2}}$. One should note that for any m in $\mathbb{M}$, $\mathbb{U}_{\underline{m}}$ is the orthogonal complement of $\mathbb{U}_{\overline{m}}$ in $\mathbb{U}$. Then, the following operators appear naturally.

Definition 18 We define the following family of projection operators on Θ. For any m_1 and m_2 in $\mathbb{M}$ denote by $\Pi_{\overline{m_1}}$, $\Pi_{\underline{m_1}}$, and $\Pi_{\underline{m_1},\overline{m_2}}$ the following projection operators:

$$\begin{aligned} \Pi_{\overline{m_1}} :&\Theta \to\Theta; \\ &[x] \mapsto(s\mapsto [x](s)\mathbb{1}_{\{s\in\mathbb{F}_{m_1}\}}) \end{aligned} \qquad \begin{aligned} \Pi_{\underline{m_1}} :&\Theta \to\Theta; \\ &[x] \mapsto(s\mapsto [x](s)\mathbb{1}_{\{s\in\mathbb{F}^c_{m_1}\}}) \end{aligned}$$

$$\begin{aligned} \Pi_{\underline{m_1},\overline{m_2}} :&\Theta \to\Theta \\ &[x] \mapsto(s\mapsto [x](s)\mathbb{1}_{\{s\in\mathbb{F}^c_{m_1}\cap\mathbb{F}^c_{m_2}\}}) \end{aligned}$$

By extension, we define, for any m in $\mathbb{M}$ the truncated Fourier transform $\mathcal{F}_{\overline{m}}$

$$\mathcal{F}_{\overline{m}} : \Xi \to \Theta; \quad [x] \mapsto [x]_{\overline{m}} = (\Pi_{\overline{m}}[x] : s \mapsto [x](s)\mathbb{1}_{\{s\in\mathbb{F}_m\}}).$$

We see that $\mathcal{F}_{\overline{m}}$ is a unitary mapping between Hilbert spaces and we should highlight that its conjugates is given by

$$\mathcal{F}^\star_{\overline{m}} : \Theta \to \Xi, \quad [x] \mapsto \sum_{s\in\mathbb{F}_m} [x](s)e_s = \Pi_{\mathbb{U}_{\overline{m}}}\mathcal{F}^\star([x]) = \mathcal{F}^\star(\Pi_{\overline{m}}[x]).$$

□

Hence, considering an inverse problem where one is interested in estimating f in Ξ when having at hand an estimate $\widehat{T(f)}$ of $T(f)$ where T is a bounded linear operator from Ξ onto itself, we will consider the family of so called projection estimators defined by $\{\widehat{f}_{\overline{m}} = \Pi_{\mathbb{U}_{\overline{m}}}(T^+\widehat{T(f)}), m \in \mathbb{M}\}$.
We will see that, often, it will be easier to approximate objects in Θ and then apply $\mathcal{F}^\star$. In this perspective we extend the definition of $\mathcal{F}$ in the following way.

Definition 19 Denote $\mathcal{L}(\Theta)$ the space of linear application from Θ onto itself. Then, for any S in $\mathcal{L}(\Xi)$, we define $[S]$ to be

$$\begin{aligned} [S] :&\mathbb{F}^2 &&\to\mathbb{K}. \\ &(s_1, s_2) &&\mapsto[S](s_1, s_2) = \langle e_{s_1}|S(e_{s_2})\rangle_\Xi \end{aligned}$$

Notice that $[S]$ defines an element of $\mathcal{L}(\Theta)$ such that, for any $[x]$ in Θ, $[S][x]$ is such that, for any s in $\mathbb{F}$, $[S][x](s)$ is given by $\sum_{s'\in\mathbb{F}}[S](s, s')[x](s')$.
In addition, we define, for any m in $\mathbb{M}$ and $[S]$ in $\mathcal{L}(\Theta)$, the operator $[S]_{\overline{m}}$ such that for any s_1 and s_2 in $\mathbb{F}$, we have $[S]_{\overline{m}}(s_1, s_2) = [S](s_1, s_2)\mathbb{1}_{\{\{s_1\in\mathbb{F}_m\}\cap\{s_2\in\mathbb{F}_m\}\}}$. It is interesting to note that for any S in $\mathcal{L}(\Xi)$ and m in $\mathbb{M}$, if we denote $S_{\overline{m}} = \Pi_{\mathbb{U}_{\overline{m}}}S\Pi_{\mathbb{U}_{\overline{m}}}$, we have $[S]_{\overline{m}} = [S_{\overline{m}}]$.
We note that the adjoint operator of $[S]$ is represented for any s_1 and s_2 in $\mathbb{F}$ by $[S]^\star(s_1, s_2) = [S^\star](s_1, s_2) = \overline{[S](s_2, s_1)}$. □

Notice that, for the operator T appearing in our model, due to Assumption 2, we have for any s and s' in $\mathbb{F}$ that $[T](s, s') = \mathbb{1}_{\{s=s'\}}\lambda(s)$. Considering the objects we just introduced, the following notations will be convenient throughout the thesis.

Notation 1 For any m in $\mathbb{M}$ let be the following objects:

$$\begin{aligned} \lambda_{\overline{m}} :& \mathbb{F} \to \mathbb{K}; & \theta^{\circ}_{\overline{m}} :& \mathbb{F} \to \mathbb{K}; & \phi_{\overline{m}} :& \mathbb{F} \to \mathbb{K}. \\ s \mapsto& \Pi_{\overline{m}}\lambda(s) & s \mapsto& \Pi_{\overline{m}}\theta^{\circ}(s) & s \mapsto& \Pi_{\overline{m}}\phi(s) = \lambda_{\overline{m}}(s)\theta^{\circ}(s) \end{aligned}$$

as well as their counterparts in Ξ

$$h_{\overline{m}} := \mathcal{F}^{-1}(\lambda_{\overline{m}})\,; \quad f_{\overline{m}} := \mathcal{F}^{-1}(\theta^{\circ}_{\overline{m}}); \quad g_{\overline{m}} := \mathcal{F}^{-1}(\phi^{\circ}_{\overline{m}}).$$

We also define their empirical counterparts which are called "projection estimators". Under Assumption 3 they take the following form:

$$\begin{aligned} \phi_{n,\overline{m}} :& \mathbb{F} \to \mathbb{K}; & \theta_{n,\overline{m}} :& \mathbb{F} \to \mathbb{K}; \\ s \mapsto& \Pi_{\overline{m}}\phi_n(s) & s \mapsto& \Pi_{\overline{m}}\theta_n(s) = \lambda^{-1}_{\overline{m}}(s)\phi_n(s) \end{aligned}$$

and their counterparts in Ξ are

$$g_{n,\overline{m}} := \mathcal{F}^{-1}(\phi_{n,\overline{m}}); \quad f_{n,\overline{m}} := \mathcal{F}^{-1}(\theta_{n,\overline{m}}).$$

On the other hand, under Assumption 4 they take the form

$$\begin{aligned} \phi_{n,\overline{m}} :& \mathbb{F} \to \mathbb{K}; & \lambda_{n_\lambda} :& \mathbb{F} \to \mathbb{K}; & \theta_{n,n_\lambda,\overline{m}} :& \mathbb{F} \to \mathbb{K}. \\ s \mapsto& \Pi_{\overline{m}}\phi_n(s) & s \mapsto& \lambda_{n_\lambda}(s) & s \mapsto& \Pi_{\overline{m}}\theta_{n,n_\lambda}(s) = \lambda^{+}_{n_\lambda}(s)\phi_{n,\overline{m}}(s) \end{aligned}$$

where $\lambda^{+}_{n_\lambda}(s) = \mathbb{1}_{\{|\lambda_{n_\lambda}(s)|^2 \geqslant n_\lambda^{-1}\}}\lambda^{-1}_{n_\lambda}(s)$, for any s in $\mathbb{F}$. Their counterparts in Ξ are

$$g_{n,\overline{m}} := \mathcal{F}^{-1}(\phi_{n,\overline{m}}); \quad f_{n,\overline{m}} := \mathcal{F}^{-1}(\theta_{n,\overline{m}}).$$

□

The family $\{\lambda_{\overline{m}}, m \in \mathbb{M}\}$ defines a regularisation as defined in Definition 16. We hence have at hand a family of estimators, called projection estimators, arising from the empirical estimators based on our data while using the dimension reduction regularisation technic.

Note that many other types of regularisations have gathered interest along the years. For example Engl et al. (1989) consider the convergence rate of Tikhonov regularisation; while Cavalier and Raimondo (2007) consider the Galerkin regularisation.

The estimation technics we will study in this thesis are deeply linked to the family of projection estimators. As one might notice, given a set of observations, the number of potential estimators for f is infinite, and it can be easily seen that most of them do not lead to a consistent estimation. Hence, we will be interested in properties which can objectively indicate if a given estimator is satisfying.

1.3.2 Decision theory

As we have seen previously, for a given model, one could chose among a variety of estimators. This choice is in general not obvious and decision theory can be used to help in this process.
To make this part more illustrative for the remaining of this script let us first introduce the following set of assumptions about the parameter space that will hold true for all of our examples.

Assumption 7 Assume that Ξ is a subset of the space of functions from $[0,1]$ to $\mathbb{C}$, equipped with the scalar product $(x,y) \mapsto \langle x|y\rangle_{L^2} = \int_{[0,1]} x(t) \cdot \overline{y(t)}\, \mathrm{d}t$. Then we consider $(e_s)_{s\in\mathbb{Z}} = ([0,1] \to \mathbb{C}, t \mapsto \exp[2 \cdot \imath \cdot \pi \cdot s \cdot t])_{s\in\mathbb{Z}}$. One can see that it is an orthonormal system in Ξ. Hence, Θ is a subset of $\mathbb{C}^{\mathbb{Z}}$ equipped with the scalar product $([x],[y]) \mapsto \langle [x]|[y]\rangle_{l^2} = \sum_{s\in\mathbb{Z}} [x](s) \cdot \overline{[y](s)}$. □

We have used, in the past sections, the distance between an estimate of an object of interest and the said object as an argument about whether one should be satisfied about the said estimate. We formalise now the criteria under which one can qualify an estimator as satisfying.

1.3.2.1 The loss function $l : (\{\mathbb{Y} \to \Xi\} \times \mathbb{Y} \times \Xi) \to \mathbb{R}_+$

this function represents the error made by using a certain estimator $\widehat{f}$ while estimating the true parameter f when the data at hand is Y.
A natural choice would be to consider a distance on Ξ, say $d : \Xi \times \Xi \to \mathbb{R}_+$ and to define $l : \{\mathbb{Y} \to \Xi\} \times \mathbb{Y} \times \Xi \to \mathbb{R}_+; \quad (\widehat{f}, Y, f) \mapsto d(\widehat{f}(Y), f)$.

Under Assumption 7 it is natural to consider an element of the family of L^p distances defined for any p in $\mathbb{R}_+$ and x and y in Ξ by $\|x-y\|_{L^p} = (\int_{[0,1]} |x(t)-y(t)|^p \, \mathrm{d}t)^{1/p}$ with the limit cases $\|x-y\|_{L^\infty} = \sup_{t\in[0,1]}\{|x(t)-y(t)|\}$ and $\|x-y\|_{L^0} = \int_{[0,1]} \mathbb{1}_{\{|x(t)-y(t)|>0\}} \, \mathrm{d}t$.
In this thesis we will only consider the quadratic loss function L^2. Notice, though, that our results could be easily generalised to the case where given a measurable function $\mathfrak{u}$ in Ξ, one considers for any x in Ξ its weighted norm $\|x\|_{L^2_{\mathfrak{u}}} = (\int_{[0,1]} |(x \star \mathfrak{u})(t)|^2 \, \mathrm{d}t)^{1/2} = (\int_{[0,1]} |(\int_{[0,1]} x(v) \cdot \mathfrak{u}(t-v) \, \mathrm{d}v)|^2 \, \mathrm{d}t)^{1/2}$ where $\star$ stands for the convolution operator on Ξ. In addition, this type of norm will nonetheless play an important role later where we consider minimax optimality over Sobolev's ellipsoids.

In order to apply decision theory, we have to assume that the object f we try to estimate belongs to the space where the loss function is finite, for which we give the following notations.

Definition 20 Let $\mathbb{L}^2$ be the subset of Ξ such that $\mathbb{L}^2 := \{x \in \Xi : \|x\|_{L^2} < \infty\}$ and in addition, for any function $\mathfrak{u}$ in Ξ and any r in $\mathbb{R}_+$ let be $\mathbb{L}^2_{\mathfrak{u}} := \{x \in \Xi : \|x\|_{L^2_{\mathfrak{u}}} < \infty\}$ and $\Xi_{\mathfrak{u}}(r) := \{x \in \Xi : \|x\|_{L^2_{\mathfrak{u}}} < r\}$. □

We have seen that we are interested in estimation methods which are based on the estimation of the Fourier transform of f, θ°. In the case of the L^2-norm, we can see that

considering the loss function on Θ is sufficient to quantify the performance on Ξ. Indeed, let be the l^2-norm on Θ defined for any $[x]$ in Θ by $\|[x]\|_{l^2} = (\sum_{s\in\mathbb{Z}} |[x](s)|^2)^{1/2}$ and the associated space $\mathcal{L}^2 = \{[x] \in \Xi : \|[x]\|_{l^2} < \infty\}$. Given a sequence $[\mathfrak{u}]$ in Θ, we can define the weighted norm which is given, for any $[x]$ in Θ by $\|[x]\|_{l^2_{[\mathfrak{u}]}} = (\sum_{s\in\mathbb{Z}} |[x](s)[\mathfrak{u}](s)|^2)^{1/2}$ and for any r in $\mathbb{R}_+$ we define the associated space $\Theta([\mathfrak{u}], r) := \{[x] \in \Theta : \|[x]\|_{l^2_{[\mathfrak{u}]}} < r\}$.
The theorem of Plancherel gives us the link between those distances, we have for any x and $\mathfrak{u}$ in Ξ and their Fourier transforms $[x]$ and $[\mathfrak{u}]$ the $\|x\|^2_{L^2_{\mathfrak{u}}} = \|[x]\|^2_{l^2_{[\mathfrak{u}]}}$.
We hence assume from now on that the parameter of interest has finite norm.

Assumption 8 The parameter of interest f is in $\mathbb{L}^2$. □

This assumption is equivalent to assuming that θ° is in $\mathcal{L}^2$.
We shall highlight that this definition has to be adapted under Assumption 4 where we obtain $l : \{\mathbb{Y}^2 \to \Xi\} \times \mathbb{Y}^2 \times \Xi \to \mathbb{R}_+; \quad (\widehat{f}, Y, \varepsilon, f) \mapsto d(\widehat{f}(Y, \varepsilon), f)$.

1.3.2.2 The risk function $(\mathcal{R}_n : (\{\mathbb{Y} \to \Theta\} \times \Theta) \to \mathbb{R}_+)_{n\in\mathbb{N}}$

One can notice the the loss function defined previously depends on the observation and, as such, is a random object that cannot be optimised over the choice of estimator.
A way to overcome this limitation is considering a so called risk function such as the expected loss function $\mathcal{R}_n(\widehat{f}, f) = \mathbb{E}\left[l(\widehat{f}, Y^n, f)\right]$ or $\mathcal{R}_{n,n_\lambda}(\widehat{f}, f, h) = \mathbb{E}\left[l(\widehat{f}, Y^n, \varepsilon^{n_\lambda}, f)\right]$ depending on the considered set of assumptions.
The following assumption, which will be verified in every model we consider allows us to obtain interesting upper bounds for the quadratic risk of projection estimators.

Assumption 9 Assume that there exist constants V_1 and V_2 in $\mathbb{R}_+\star$ such that, for any s in $\mathbb{F}$, we have $V_1 \leqslant \mathbb{V}[\langle Y_0|e_s\rangle_\Xi] \leqslant V_2$. In addition assume that there exist constants V_3, V_4, and $\mathcal{C}_4$ such that $V_3 \leqslant \mathbb{V}[\langle \varepsilon_0|e_s\rangle_\Xi] \leqslant V_4$ and $n_\lambda^2\, \mathbb{E}|\lambda(s) - \lambda_{n_\lambda}(s)|^4 \leqslant \mathcal{C}_4$. □

This hypothesis allows us to show the following result.

Lemma 1.3.1.
If Assumption 9 holds true, then (i) $\mathbb{E}|\lambda(s)\lambda^+_{n_\lambda}(s)|^2 \leqslant 2V_4 + 1$; *(ii)* $\mathbb{P}(|\lambda^+_{n_\lambda}(s)|^2 < 1/n_\lambda) \leqslant 4V_4(1 \wedge \Lambda(s)/n_\lambda)$, *(iii)* $\mathbb{E}|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\lambda^+_{n_\lambda}(s)|^2 \leqslant 2(\mathcal{C}_4 + V_4)(1 \wedge \Lambda(s)/n_\lambda)$.

Proof of Lemma 1.3.1
Since $n_\lambda\, \mathbb{E}|\lambda(s) - \lambda_{n_\lambda}(s)|^2 = \mathbb{V}[\langle \varepsilon_0|e_s\rangle_\Xi] \leqslant V_4$ we obtain (i) as follows

$$\begin{aligned}\mathbb{E}|\lambda(s)\lambda^+_{n_\lambda}(s)|^2 &\leqslant 2\,\mathbb{E}\{|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\lambda^+_{n_\lambda}(s)|^2 + \mathbb{1}_{\{|\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda\}}\} \\ &\leqslant 2(n_\lambda\, \mathbb{E}(\lambda(s) - \lambda_{n_\lambda}(s))^2 + 1) \leqslant 2V_4 + 1.\end{aligned}$$

Consider (ii). Trivially, for any $s \in \mathbb{N}$ we have $\mathbb{P}(|\lambda_{n_\lambda}(s)|^2 < 1/n_\lambda) \leqslant 1$. If $1 \leqslant 4V_4 n_\lambda^{-1}|\lambda(s)|^{-2}) = 4V_4 n_\lambda^{-1}\Lambda(s)$, then obviously $\mathbb{P}(|\lambda_{n_\lambda}(s)|^2 < n_\lambda^{-1}) \leqslant \min(1, 4V_4 n_\lambda^{-1}\Lambda(s))$. Otherwise, we have $n_\lambda^{-1} < |\lambda(s)|^2/(4V_4)$ and hence using Tchebychev's inequaltiy,

$$\begin{aligned}\mathbb{P}(|\lambda_{n_\lambda}(s)|^2 < n_\lambda^{-1}) &\leqslant \mathbb{P}(|\lambda_{n_\lambda}(s) - \lambda(s)| > |\lambda(s)|/(2\sqrt{V_4})) \leqslant 4\Lambda(s)\,\mathbb{E}|\lambda(s) - \lambda_{n_\lambda}(s)|^2 \\ &\leqslant 4V_4 n_\lambda^{-1}\Lambda(s) = \min(1, 4V_4 n_\lambda^{-1}\Lambda(s))\end{aligned}$$

where we have used again that $n_\lambda\,\mathbb{E}|\lambda(s)-\lambda_{n_\lambda}(s)|^2 \leqslant 1$. Combining both cases we obtain (ii).

Consider (iii). Due to Assumption 9 there is a numerical constant $\mathcal{C}_4$ such that $n_\lambda^2\,\mathbb{E}\,|\lambda(s)-\lambda_{n_\lambda}(s)|^4 \leqslant \mathcal{C}_4$, which in turn implies

$$\mathbb{E}|\lambda(s)-\lambda_{n_\lambda}(s)|^2|\lambda^+_{n_\lambda}(s)|^2 \leqslant \mathbb{E}\left\{|\lambda(s)-\lambda_{n_\lambda}(s)|^2|\lambda^+_{n_\lambda}(s)|^2 2\Big[\frac{|\lambda(s)-\lambda_{n_\lambda}(s)|^2}{|\lambda(s)|^2}+\frac{|\lambda_{n_\lambda}(s)|^2}{|\lambda(s)|^2}\Big]\right\}$$
$$\leqslant \frac{2n_\lambda\,\mathbb{E}|\lambda(s)-\lambda_{n_\lambda}(s)|^4}{|\lambda(s)|^2}+\frac{2\,\mathbb{E}|\lambda(s)-\lambda_{n_\lambda}(s)|^2}{|\lambda(s)|^2} \leqslant 2(\mathcal{C}_4+V_4)n_\lambda^{-1}\Lambda(s).$$

Combining the last bound and $\mathbb{E}|\lambda(s)-\lambda_{n_\lambda}(s)|^2|\lambda^+_{n_\lambda}(s)|^2 \leqslant n_\lambda\,\mathbb{E}|\lambda(s)-\lambda_{n_\lambda}(s)|^2 \leqslant V_4$ implies (iii), which completes the proof. □

In addition we will use the following notations.

NOTATION 2 For any m in $\mathbb{M}$; s in $\mathbb{F}$; and θ in Θ, let be the following quantities:

$$\mathfrak{b}_m^2(\theta) := \|\theta_{\underline{0}}\|_{l^2}^{-2}\|\theta_{\underline{m}}\|^2; \quad \Lambda(s) = |\lambda^{-1}(s)|^2;$$
$$\Lambda_\circ(m) = m^{-1}\sum\nolimits_{0<s\leqslant m}\Lambda(s); \quad \Lambda_+(m) := \max_{s\in\mathbb{F}_m}\{\Lambda(s)\}.$$

□

Notice that, if $\mathbb{F}=\mathbb{Z}$, then $\sum_{s\in\mathbb{F}_m}\Lambda(s) = 2m\Lambda_\circ(m)+\Lambda(0)$ and if $\mathbb{F}=\mathbb{N}^\star$ then $\sum_{s\in\mathbb{F}_m}\Lambda(s) = m\Lambda_\circ(m)$. So in both case we will write $\sum_{s\in\mathbb{F}_m}\Lambda(s) = \mathcal{C}m\Lambda_\circ(m)$.

EXAMPLE 1.3.1 PROJECTION ESTIMATOR

If one considers a projection estimator, as in notation 1, one can carry the following computations out for any m in $\mathbb{M}$,

$$\mathcal{R}_n(\theta_{n,\overline{m}},\theta,\Lambda) = \mathbb{E}\left[\|\theta_{n,\overline{m}}-\theta\|_{l^2}^2\right] = \sum\nolimits_{s\in\mathbb{F}}\mathbb{V}\left[\theta_{n,\overline{m}}(s)\right]+|\,\mathbb{E}\left[\theta_{n,\overline{m}}(s)\right]-\theta(s)|^2.$$

Them, under Assumption 9, the quadratic risk can be simplified, depending on the set of assumptions we accept:

- *under Assumption 3 and Assumption 5*

$$\mathcal{R}_n(\theta_{n,\overline{m}},\theta,\Lambda) = n^{-1}\sum\nolimits_{s\in\mathbb{F}_m}\Lambda(s)\,\mathbb{V}\left[\langle Y_0|e_s\rangle_{L^2}\right]+\|\theta_{\underline{0}}\|_{l^2}^2\mathfrak{b}_m^2(\theta)$$
$$\leqslant n^{-1}V_2\mathcal{C}m\Lambda_\circ(s)+\|\theta_{\underline{0}}\|_{l^2}^2\mathfrak{b}_m^2(\theta) \leqslant (V_2\mathcal{C}+\|\theta^\circ_{\underline{0}}\|_{l^2}^2)[n^{-1}m\Lambda_\circ(m)\vee\mathfrak{b}_m^2(\theta^\circ)];$$

but also

$$\mathcal{R}_n(\theta_{n,\overline{m}},\theta,\Lambda) \geqslant n^{-1}V_1\mathcal{C}m\Lambda_\circ(s)+\|\theta_{\underline{0}}\|_{l^2}^2\mathfrak{b}_m^2(\theta) \geqslant (V_1\mathcal{C}\vee\|\theta^\circ_{\underline{0}}\|_{l^2}^2)[n^{-1}m\Lambda_\circ(m)\vee\mathfrak{b}_m^2(\theta^\circ)];$$

- *under Assumption 3 and Assumption 6, using Lemma 1.2.1,*

$$\mathcal{R}_n(\theta_{n,\overline{m}}, \theta, \Lambda) = n^{-2} \sum_{s \in \mathbb{F}_m} \Lambda(s)\, \mathbb{V}\left[\sum_{p=1}^{n} \langle Y_p | e_s \rangle_{L^2}\right] + \|\theta_{\underline{0}}\|_{l^2}^2 \mathfrak{b}_m^2(\theta)$$
$$\leqslant (\mathcal{C}(1 + 4 \sum_{p=1}^{\infty} \beta(Y_0, Y_p)) + \|\theta_{\underline{0}}\|_{l^2}^2)[n^{-1} m \Lambda_\circ(s) \vee \mathfrak{b}_m^2(\theta)];$$

- *under Assumption 4 and Assumption 5, start by noticing that, as for any s in $\mathbb{F}$, we have $(1 \wedge \Lambda(s)) \leqslant 1$ and that θ° is square summable, we have, $\sum_{s \in \mathbb{F}} |\theta^\circ(s)|^2 (1 \wedge n_\lambda^{-1} \Lambda(s)) < \infty$. Hence, using Lemma 1.3.1, we may write,*

$$\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m}}, \theta, \Lambda) = \sum_{s \in \mathbb{F}_m} \Lambda(s) \left(\mathbb{V}\left[\phi_n(s)\right] \mathbb{E}\left[|\lambda_{n_\lambda}^+(s) \lambda(s)|^2\right]\right) + \|\theta_{\underline{0}}\|_{l^2}^2 \mathfrak{b}_m^2(\theta)$$
$$+ \sum_{s \in \mathbb{F}_m} |\theta(s)|^2\, \mathbb{E}\left[\left|\lambda_{n_\lambda}^+(s)\right|^2 |\lambda(s) - \lambda_{n_\lambda}(s)|^2\right] + \sum_{s \in \mathbb{F}_m} |\theta(s)|^2\, \mathbb{P}(\{|\lambda_{n_\lambda}(s)|^2 < n_\lambda^{-1}\})$$
$$\leqslant (V_2 \mathcal{C} + \|\theta_{\underline{0}}\|_{l^2}^2)[n^{-1} m \Lambda_\circ(m) \vee \mathfrak{b}_m^2(\theta)] + 2\mathcal{C}(\mathcal{C}_4 + 3V_4) \sum_{s \in \mathbb{F}} |\theta(s)|^2 (1 \wedge n_\lambda^{-1} \Lambda(s)).$$

□

Notation 3 In particular, we denote in the following way the risk for projection estimators:

$$\mathcal{R}_n^m(\theta^\circ, \Lambda) := [n^{-1} m \Lambda_\circ(m) \vee \mathfrak{b}_m^2(\theta^\circ)]; \quad \mathcal{R}_{n_\lambda}^\dagger(\theta^\circ, \Lambda) := \sum_{s \in \mathbb{F}} |\theta(s)|^2 (1 \wedge n_\lambda^{-1} \Lambda(s)).$$

The risk function hence allows us to quantify the performance of an estimator independently of the random observation. Alternatively, one can consider the probability to exceed a certain loss.

Definition 21 We define the sequence of functions

$$\mathfrak{R}_n : (\mathbb{Y} \to \Xi) \times \Xi \times \mathbb{R}_+ \to \mathbb{R}_+; \quad (\widehat{f}, f, a) \mapsto \mathbb{P}_f^n\left(l(\widehat{f}, Y, f) \geqslant a\right).$$

□

In general, one is interested in the asymptotic behaviour of $\mathcal{R}$ or $\mathfrak{R}$ (and then replacing a by a sequence $(a_n)_{n \in \mathbb{N}}$) when n tends to infinity. In particular, for a given estimator $\widehat{f}$ and a fixed value f of the parameter of interest, the sequence $\mathcal{R}_n(\widehat{f}, f)$ is called convergence rate of $\widehat{f}$ at f and if $\mathfrak{R}_n(\widehat{f}, f, a_n)$ tends to 0 as n tends to infinity, a_n is called speed of convergence in probability of $\widehat{f}$ at f. If this sequence tends to zero, the estimator is called consistent.

While it is technically feasible to minimise the risk function over $\widehat{f}$ for each f, the result will be discountenancing as the minimisers will invariably be functions almost surely equal to f itself which brilliantly yields a loss function equal to 0, independently of the observation and hence a risk function equal to 0. Our goal being to estimate f, it is obvious that such an estimator is not at hand.
We are interested in this thesis in two formulations of optimality which allow to overcome this limitation.

1.3.2.3 Oracle optimality

Consider $\mathcal{E}$, a family of estimators and a risk function $\mathcal{R}$.

Definition 22 A sequence of functions $(\mathcal{R}_{\mathcal{E},n} : \Xi \to \mathbb{R}_+)_{n\in\mathbb{N}}$ is called oracle risk for the family of estimators $\mathcal{E}$ if there exist a constant C in $[1, \infty[$ such that, for any f in Ξ, and all n, we have: $\mathcal{R}_{\mathcal{E},n}(f) \leqslant C \cdot \inf_{\widehat{f}\in\mathcal{E}} \mathcal{R}_n(\widehat{f}, f)$, or, depending on the considered set of assumptions, $\mathcal{R}_{\mathcal{E},n,n_\lambda}(f,h) \leqslant C \cdot \inf_{\widehat{f}\in\mathcal{E}} \mathcal{R}_{n,n_\lambda}(\widehat{f}, f, h)$. □

Definition 23 A sequence of functions $\mathcal{R}^\circ_{\mathcal{E},n} : \Xi \to \mathbb{R}_+$ is called exact oracle convergence rate for the family of estimators $\mathcal{E}$ if, in addition to being an oracle convergence rate, there exists an element $\widehat{f}$ of $\mathcal{E}$ such that for any f in Ξ and n in $\mathbb{N}$ we have: $\mathcal{R}^\circ_{\mathcal{E},n}(f) \geqslant C^{-1} \cdot \mathcal{R}_n(\widehat{f}, f)$ or $\mathcal{R}^\circ_{\mathcal{E},n,n_\lambda}(f,h) \geqslant C^{-1} \cdot \mathcal{R}_{n,n_\lambda}(\widehat{f}, f, h)$ depending on the type of data at hand. An estimator such as $\widehat{f}$ is called oracle optimal. □

We see that those definitions are "up to a constant" and we will in general be more interested in the asymptotic rate as n and/or n_λ tend to infinity and we hence introduce the following notations.

Notation 4 Let be $(a_n)_{n\in\mathbb{N}}$ a sequence of elements of $\mathbb{K}$. We define the sets $\mathfrak{o}_n(a) := \{b \in \mathbb{K}^{\mathbb{N}} : \lim_{n\to\infty} |b_n/a_n| = 0\}$; and $\mathcal{O}_n(a) := \{b \in \mathbb{K}^{\mathbb{N}} : \exists C \in \mathbb{R}_+ \lim_{n\to\infty} |b_n/a_n| \leqslant C\}$. If $a \in \mathcal{O}(b)$ and $b \in \mathcal{O}(a)$ then we denote $a \approx b$.
On the other hand we also define the sets $\mathfrak{o}_{\mathbb{P}}(a)$ and $\mathcal{O}_{\mathbb{P}}(a)$ as the sets of sequences of probability distributions $\mathbb{P}_n$ on $\mathbb{K}$ such that, if $(X_n)_{n\in\mathbb{N}}$ is a sequence of $\mathbb{K}$-valued random variables verifying $X_n \sim \mathbb{P}_n$, then we have,

$$\mathbb{P}_n \in \mathfrak{o}_{\mathbb{P}}(a) \iff \forall \varepsilon \in \mathbb{R}^\star_+, \quad \lim_{n\to\infty} \mathbb{P}(|X_n/a_n| \geqslant \varepsilon) = 0$$

$$\mathbb{P}_n \in \mathcal{O}_{\mathbb{P}}(a) \iff \forall \varepsilon \in \mathbb{R}^\star_+, \exists M \in \mathbb{R}_+, N \in \mathbb{N} : \quad \forall n > N, \mathbb{P}(|X_n/a_n| \geqslant M) \leqslant \varepsilon$$

□

In particular, throughout this thesis, we shall distinguish the following two cases for θ°, respectively called parametric and non-parametric which commonly lead to very different behaviour of the optimal rates:

(p) there exist a finite K of $\mathbb{F}$ such that, for any K' smaller than K, $\mathfrak{b}^2_{K'}(\theta^\circ) > 0$ and $\mathfrak{b}^2_K(\theta^\circ) = 0$;

(np) for all finite K in $\mathbb{F}$, $\mathfrak{b}_K(\theta^\circ) > 0$.

Note that the Fourier series expansion of the function of interest f is, in case (p), *finite*, i.e., $f = \sum_{s\in\mathbb{F}_K} \theta^\circ(s) e_s$ for some finite K in $\mathbb{N}$ while in the opposite case (np), it is *infinite*, i.e., not finite.

Numerical discussion 1.3.1.
The upper bounds we give will be discussed in such "numerical discussions" where we consider the following typical behaviours of θ° and λ and give an equivalent to the upper bound in terms of an explicit function of n.

Regarding the operator eigen-values λ, we consider the following two cases, respectively called ordinary smooth and super-smooth:

(o) there exists a strictly positive real number a such that $\Lambda(m) \approx m^{2a}$, then $m\Lambda_{\circ}(m) \approx m^{2a+1}$ and $\Lambda_{+}(m) \approx m^{2a}$;

(s) there exists a strictly positive real number a such that $\Lambda(m) \approx \exp(m^{2a})$, then $m\Lambda_{\circ}(m) \approx m^{-(1-2a)_{+}} \exp(m^{2a})$ and $\Lambda_{+}(m) \approx \exp(m^{2a})$.

For the parameter of interest θ°, the behaviours of its tails i.e., $\left(\mathfrak{b}_m^2(\theta^{\circ})\right)_{m\in\mathbb{F}} = \|\theta^{\circ}_{\underline{0}}\|_{l^2}^{-2}\|\Pi_{\underline{m}}\theta^{\circ}\|_{l^2}^2$ will also be of interest. We distinguish the cases **(p)** and **(np)**, and with **(np)** distinguish the super smooth and ordinary smooth for the parameter of interest.

(o) there exists a strictly positive real number p such that $|\theta^{\circ}(s)|^2 \approx s^{-2p-1}$, in this case, we have $\mathfrak{b}_m^2(\theta^{\circ}) \approx m^{-2p}$;

(s) there exists a strictly positive real number p such that $|\theta^{\circ}(s)|^2 \approx s^{2p-1}\exp[-s^{2p}]$, and then we have $\mathfrak{b}_m^2(\theta^{\circ}) \approx \exp(-m^{2p})$.

We consider the following situations: in the cases **[p-o]** and **[p-s]** the parameter of interest has a finite representation **(p)** and the operator is either ordinary smooth **(o)** or super smooth **(s)**. In the cases **[o-o]** and **[o-s]** the parameter of interest is ordinary smooth **(o)** and the operator is either ordinary smooth **(o)** or super smooth **(s)**. Case **[s-o]** is the opposite of case **[o-s]**. □

While the names given here to the typical cases may seem arbitrary, we shall justify them through the examples treated in this thesis where the decaying rate of θ° and λ respectively can be interpreted in terms of function smoothness.
The particular interest for these different cases will also appear natural as the behaviour of the optimal rate will be considerably different in our examples; moreover, this phenomenon is observed in many statistical models, also outside of our field of interest.
We carry on with the projection estimators example.

Known operator

The bound we derived in notation 3 depends on the dimension parameter m and hence by selecting an optimal value they will be minimised, which we formulate next. For a sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers with minimal value in a set $A \subset \mathbb{N}$ we set $\arg\min\{a_n, n \in A\} := \min\{m \in A : a_m \leqslant a_n,\ \forall n \in A\}$. For all $n \in \mathbb{N}$ we define

$$\mathcal{R}_n^m(\theta^{\circ},\Lambda) := [\mathfrak{b}_m^2(\theta^{\circ}) \vee m\Lambda_{\circ}(m)n^{-1}] := \max\left(\mathfrak{b}_m^2(\theta^{\circ}), m\Lambda_{\circ}(m)n^{-1}\right),$$
$$m_n^{\circ} := m_n^{\circ}(\theta^{\circ},\Lambda) := \arg\min\{\mathcal{R}_n^m(\theta^{\circ},\Lambda), m \in \mathbb{N}\} \quad \text{and}$$
$$\mathcal{R}_n^{\circ}(\theta^{\circ},\Lambda) := \mathcal{R}_n^{m_n^{\circ}}(\theta^{\circ},\Lambda) = \min\{\mathcal{R}_n^m(\theta^{\circ},\Lambda), m \in \mathbb{N}\}. \quad (1.1)$$

Consequently, the rate $(\mathcal{R}_n^{\circ}(\theta^{\circ},\Lambda))_{n\in\mathbb{N}}$, the dimension parameters $(m_n^{\circ})_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m_n^{\circ}}})_{n\in\mathbb{N}}$, respectively, is an oracle rate, an oracle dimension and oracle optimal estimator (up to a constant).

Remark 1.3.1 *We shall emphasise that $\mathcal{R}_n^{\circ}(\theta^{\circ},\Lambda) \geqslant n^{-1}$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty}\mathcal{R}_n^{\circ}(\theta^{\circ},\Lambda) = 0$. Observe that for all $\delta > 0$ there exists $m_\delta \in \mathbb{N}$ and $n_\delta \in \mathbb{N}$ such that for all $n \geqslant n_\delta$ holds $\mathfrak{b}_{m_\delta}^2(\theta^{\circ}) \leqslant \delta$ and $m_\delta\Lambda_{\circ}(m_\delta)n^{-1} \leqslant \delta$, and whence*

$\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \leqslant \mathcal{R}_n^{m_\delta}(\theta^\circ,\Lambda) \leqslant \delta$. Moreover, we have $m_n^\circ \in [\![1,n]\!]$. Indeed, by construction holds $\mathfrak{b}_n^2(\theta^\circ) \leqslant 1 < (n+1)n^{-1} \leqslant (n+1)\Lambda_\circ(n+1)n^{-1}$, and hence $\mathcal{R}_n^n(\theta^\circ,\Lambda) < \mathcal{R}_n^m(\theta^\circ,\Lambda)$ for all $m \in [\![n+1,\infty[\![$ which in turn implies the claim $m_n^\circ \in [\![1,n]\!]$. Obviously, it follows thus $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \min\{\mathcal{R}_n^m(\theta^\circ,\Lambda), m \in [\![1,n]\!]\}$ for all $n \in \mathbb{N}$. We shall use those elementary findings in the sequel without further reference. The sequence $\mathcal{R}_n^\circ(\theta,\lambda)$ is then an exact oracle convergence rate and the projection estimator $\theta_{n,\overline{m_n^\circ}}$ is an oracle optimal estimator. □

Remark 1.3.2 *In case **(p)**, the oracle rate is parametric, that is $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx n^{-1}$. More precisely, if $\theta^\circ = 0$ then for each $m \in \mathbb{N}$, $\mathbb{E}\|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2 = \mathcal{C}m\Lambda_\circ(m)n^{-1}$, and hence $m_n^\circ = 1$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \Lambda_\circ(1)n^{-1} \sim n^{-1}$. Otherwise if there is $K \in \mathbb{N}$ with $\mathfrak{b}_{K-1}(\theta^\circ) > 0$ and $\mathfrak{b}_K(\theta^\circ) = 0$, then setting $n_{\theta^\circ} := \frac{K\Lambda_\circ(K)}{\mathfrak{b}_{K-1}^2(\theta^\circ)}$, for all $n \geqslant n_{\theta^\circ}$ holds $\mathfrak{b}_{K-1}^2(\theta^\circ) > K\Lambda_\circ(K)n^{-1}$, and hence $m_n^\circ = K$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = K\Lambda_\circ(K)n^{-1} \sim n^{-1}$. On the other hand side, in case **(np)** the oracle rate is non-parametric, more precisely, it holds $\lim_{n\to\infty} n\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \infty$. Indeed, since $\mathfrak{b}_{m_n^\circ}^2(\theta^\circ) \leqslant \mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \mathcal{R}_n^\circ(\theta^\circ,\Lambda) \in \mathfrak{o}_n(1)$ follows $m_n^\circ \to \infty$ and hence $m_n^\circ\Lambda_\circ(m_n^\circ) \to \infty$ which implies the claim because $n\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \geqslant m_n^\circ\Lambda_\circ(m_n^\circ)$.*

Numerical discussion 1.3.2.
Let us illustrate the rates obtained in the case **(np)**.

[o-o] $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx (m_n^\circ)^{-2p} \approx (m_n^\circ)^{2a+1}n^{-1}$, and hence, $m_n^\circ \approx n^{1/(2p+2a+1)}$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx n^{-2p/(2p+2a+1)}$

[o-s] $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx (m_n^\circ)^{-2p} \approx (m_n^\circ)^{-(1-2a)_+}\exp((m_n^\circ)^{2a})n^{-1}$, and hence, $m_n^\circ \approx (\log n - \frac{2p-(1-2a)_+}{2a}\log\log n)^{1/(2a)}$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx (\log n)^{-p/a}$.

[s-o] $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx \exp(-(m_n^\circ)^{2p}) \approx (m_n^\circ)^{2a+1}n^{-1}$, and hence, $m_n^\circ \approx (\log n - \frac{2a+1}{2p}\log\log n)^{1/(2p)}$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx (\log n)^{(2a+1)/(2p)}n^{-1}$.

□

Unknown operator
Let us remind that we have

$$\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m_n^\circ}}) \leqslant (V_2\mathcal{C} + \|\theta_{\underline{0}}^\circ\|_{l^2}^2)\mathcal{R}_n^\circ(\theta^\circ,\Lambda) + 2\mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)$$

We note that $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 = 0$ implies $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) = 0$, while for $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 > 0$ holds $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) \geqslant \sum_{s:\Lambda(s)>n_\lambda}|\theta^\circ(s)|^2 + n_\lambda^{-1}\sum_{s:\Lambda(s)\leqslant n_\lambda}|\theta^\circ(s)|^2 \geqslant n_\lambda^{-1}\sum_{s\in\mathbb{N}}|\theta^\circ(s)|^2 = \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 n_\lambda^{-1}$, thereby whenever $\theta^\circ \neq 0$ any additional term of order $n^{-1} + n_\lambda^{-1}$ is negligible with respect to the rate $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) + \mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)$, since $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \geqslant n^{-1}$, which we will use below without further reference. We shall emphasise that in case $n = n_\lambda$ it holds

$$\begin{aligned}\mathcal{R}_n^\dagger(\theta^\circ,\Lambda) &= \sum_{s\in\mathbb{F}_{m_n^\circ}}|\theta^\circ(s)|^2[1\wedge n^{-1}\Lambda(s)] + \sum_{s\in\mathbb{F}_{m_n^\circ}^c}|\theta^\circ(s)|^2[1\wedge n^{-1}\Lambda(s)]\\ &\leqslant \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 n^{-1}m_n^\circ\Lambda_\circ(m_n^\circ) + \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_n^\circ}^2(\theta^\circ) \leqslant \|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathcal{R}_n^{m_n^\circ}(\theta^\circ,\Lambda) \quad (1.2)\end{aligned}$$

which in turn implies $\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m_n^\circ}}) \leqslant (V_2\mathcal{C} + (1+2\mathcal{C})\|\theta_{\underline{0}}^\circ\|_{l^2}^2)\mathcal{R}_n^\circ(\theta^\circ,\Lambda)$. In other words, the estimation of the unknown operator T is negligible whenever $n \leqslant n_\lambda$.

REMARK 1.3.3 *We note that in case (p)* $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \leqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\Lambda_+(K)n_\lambda^{-1}$ *and hence*

$$\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m^\circ_n}}, \theta^\circ, \Lambda) \leqslant \mathcal{C}\{[1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2}]\{K\Lambda_\circ(K)n^{-1} + \Lambda_+(K)n_\lambda^{-1}\} \tag{1.3}$$

for all $n_\lambda \in \mathbb{N}$ *and* $n \geqslant n_{\theta^\circ}$ *with* n_{θ° *as in Remark 1.3.1. In other words the rate is parametric in both the* ε*-sample size* n_λ *and the* Y*-sample size* n*. Thereby, the additional estimation of the operator is negligible whenever* $n_\lambda \geqslant n$*. In the opposite case (np), it is obviously of interest to characterise the minimal size* n_λ *of the additional sample from* ε *needed to attain the same rate as in case of a known operator. Thus, in the next illustration we let the* ε*-sample size depend on the* Y*-sample size* n *as well.* □

Let us now briefly illustrate the rates we already defined by stating the order of $m^\circ_n(\theta^\circ, \Lambda)$ and $\mathcal{R}^\circ_n(\theta^\circ, \Lambda)$ for the cases introduced in Num. discussion 1.3.1.

NUMERICAL DISCUSSION 1.3.3.**[o-o]** For $p > a$ holds $\sum_{s\in\mathbb{N}} |\theta^\circ(s)|^2\Lambda(s) < \infty$, and hence $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \approx n_\lambda^{-1}$, while for $p = a$ and $p < a$ holds $\sum_{s=1}^{m} |\theta^\circ(s)|^2\Lambda(s) \approx \log(m)$ and $\sum_{s=1}^{m} |\theta^\circ(s)|^2\Lambda(s) \approx m^{2(a-p)}$, respectively. For $p \leqslant a$ with $m_{n_\lambda} := \lfloor n_\lambda^{1/2a} \rfloor$ it follows $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \approx n_\lambda^{-1} \sum_{s\in[\![1,m_{n_\lambda}]\!]} \Lambda((s))|\theta^\circ(s)|^2 + \mathfrak{b}_{m_{n_\lambda}}(\theta^\circ)^2$, and thereby, $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \approx \log(n_\lambda)n_\lambda^{-1}$ for $p = a$, while $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \approx n_\lambda^{-p/a}$ for $p < a$.

[o-s] Since $\sum_{s=1}^{m} |\theta^\circ(s)|^2\Lambda(s) \approx m^{-2p-1}\Lambda(m)$ the decomposition in **[o-o]** with $m_{n_\lambda} := \lfloor (\log n_\lambda)^{1/(2a)} \rfloor$ implies $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \approx (\log n_\lambda)^{-p/a}$.

[s-o] Since $\sum_{s\in\mathbb{N}} |\theta^\circ(s)|^2\Lambda(s) < \infty$ it follows $\mathcal{R}^{\dagger}_{n_\lambda}(\theta^\circ, \Lambda) \approx n_\lambda^{-1}$.

□

We see that, given a family of estimators, oracle optimality defines the best element of this family. However, this requires to restrict ourselves to a family of estimator.

1.3.2.4 Minimax optimality

An alternative to oracle optimality is minimax optimality.

DEFINITION 24 Considering a subset $\widetilde{\Xi}$ of Ξ, and an estimator $\widetilde{f}$, we call "maximal convergence rate of $\widetilde{f}$ over $\widetilde{\Xi}$" the sequence indexed by n defined by $\mathcal{R}_n(\widetilde{f}, \widetilde{\Xi}, \Lambda) := \sup_{f\in\widetilde{\Xi}} \mathcal{R}_n(\widetilde{f}, f)$.

Alternatively, if the operator is unknown, we denote $\widetilde{\Xi}$ and $\widetilde{\mathcal{L}(\Xi)}$ two subsets, respectively of Ξ and $\mathcal{L}(\Xi)$ and we have $\mathcal{R}_{n,n_\lambda}(\widetilde{f}, \widetilde{\Xi}, \widetilde{\mathcal{L}(\Xi)}) := \sup_{T\in\widetilde{\mathcal{L}(\Xi)}} \sup_{f\in\widetilde{\Xi}} \mathcal{R}_{n,n_\lambda}(\widetilde{f}, f, T)$. □

We see here that the maximal convergence rate of an estimator corresponds to its worst case scenario over a set of true parameters. The idea will be to find an estimator with the best worst case scenario.

DEFINITION 25 Considering a subset $\widetilde{\Xi}$ of Ξ, a sequence $\mathcal{R}^\star_n(\widetilde{\Xi}, \Lambda)$ is called minimax convergence rate if there exist a constant C greater than 1 such that, for any n in $\mathbb{N}$ $\mathcal{R}^\star_n(\widetilde{\Xi}, \Lambda) \leqslant C \cdot \inf_{\widetilde{f}\in\{\mathbb{Y}\to\Xi\}} \mathcal{R}_n(\widetilde{f}, \widetilde{\Xi}, \Lambda)$ where the infimum is taken over all possible estimator.

Moreover, $\mathcal{R}^\star_n(\widetilde{\Xi},\Lambda)$ is called minimax optimal convergence rate if there exists some estimator $\widehat{f}$ such that $\mathcal{R}^\star_n(\widetilde{\Xi},\Lambda) \geqslant C^{-1}\cdot\mathcal{R}_n(\widehat{f},\widetilde{\Xi},\Lambda)$. An estimator such as $\widehat{f}$ is called minimax optimal. □

In this definition, be aware that the infimum is taken over all possible estimator of f.
An example of space which we use in this thesis as $\widetilde{\Theta}$ are Sobolev's ellipsoids which we already introduced informally previously.

Definition 26 Given a constant r in $\mathbb{R}_+$, and a positive, decreasing sequence of numbers smaller than 1, $(\mathfrak{a}(s))_{s\in\mathbb{F}}$, we define the Sobolev's ellipsoid $\Theta(\mathfrak{a},r)$ by $\Theta(\mathfrak{a},r) := \{\theta\in\Theta : \|\theta\|_{\mathfrak{a}} \leqslant r\}$. □

Those ellipsoid are interesting as they can directly be related to classes of regularity for the counterpart space Ξ.
We now carry on with the projection estimator example.
Known operator
While considering projection estimators, in the case where the operator is known, we may emphasise that for all $m\in\mathbb{N}^\star$ and any $\theta^\circ\in\Theta(\mathfrak{a},r)$, $\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ) = \|\theta^\circ_{\underline{m}}\|^2_{l^2} = \sum_{|s|>m}(\mathfrak{a}(s)^2/\mathfrak{a}(s)^2)\theta^\circ(s)^2 \leqslant \mathfrak{a}(m)^2\|\theta^\circ_{\underline{m}}\|^2_{1/\mathfrak{a}} \leqslant \mathfrak{a}(m)^2r^2$ which we use in the sequel without further reference. It follows for all $m,n\in\mathbb{N}$ that

$$\mathcal{R}_n\big(\theta_{n,\overline{m}},\Theta(\mathfrak{a},r),\Lambda\big) := \sup\big\{\mathcal{R}_n\big(\theta_{n,\overline{m}},\theta^\circ,\Lambda\big),\theta^\circ\in\Theta(\mathfrak{a},r)\big\} \leqslant (2+r^2)\max\big(\mathfrak{a}(m)^2, m\Lambda_\circ(m)n^{-1}\big). \quad (1.4)$$

The upper bound in the last display depends on the dimension parameter m and hence by choosing an optimal value $m^\star_n$ the upper bound will be minimised which we formulate next. For all $n\in\mathbb{N}$ we define

$$\begin{aligned}\mathcal{R}^m_n(\mathfrak{a},\Lambda) &:= [\mathfrak{a}(m)^2 \vee m\Lambda_\circ(m)n^{-1}] := \max\big(\mathfrak{a}(m)^2, m\Lambda_\circ(m)n^{-1}\big),\\ m^\star_n(\mathfrak{a}) &:= m^\star_n(\mathfrak{a},\Lambda) := \arg\min\{\mathcal{R}^m_n(\mathfrak{a},\Lambda), m\in\mathbb{N}\} \quad \text{and}\\ \mathcal{R}^\star_n(\mathfrak{a},\Lambda) &:= \mathcal{R}^{m^\star_n(\mathfrak{a})}_n(\mathfrak{a},\Lambda) = \min\{\mathcal{R}^m_n(\mathfrak{a},\Lambda), m\in\mathbb{N}\}. \quad (1.5)\end{aligned}$$

From (1.4) we deduce that $\mathcal{R}_n\big(\theta_{n,\overline{m^\star_n(\mathfrak{a})}},\Theta(\mathfrak{a},r),\Lambda\big) \leqslant (2+r^2)\mathcal{R}^\star_n(\mathfrak{a},\Lambda)$ for all $n\in\mathbb{N}$. On the other hand side, for example, Johannes and Schwarz (2013a) have shown that $\inf_{\widetilde{\theta}}\mathcal{R}_n\big(\widetilde{\theta},\Theta(\mathfrak{a},r),\Lambda\big)$, where the infimum is taken over all possible estimators $\widetilde{\theta}$ of θ°, is up to a constant bounded from below by $\mathcal{R}^\star_n(\mathfrak{a},\Lambda)$. Consequently, the rate $(\mathcal{R}^\star_n(\mathfrak{a},\Lambda))_{n\in\mathbb{N}}$, the dimension parameters $(m^\star_n(\mathfrak{a}))_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m^\star_n(\mathfrak{a})}})_{n\in\mathbb{N}}$, respectively, is a minimax rate, a minimax dimension and minimax optimal (up to a constant).

Remark 1.3.4 *By construction it holds $\mathcal{R}^\star_n(\mathfrak{a},\Lambda) \geqslant n^{-1}$ for all $n\in\mathbb{N}$. The following statements can be shown using the same arguments as in Remark 1.3.1 by exploiting that the sequence $\mathfrak{a}$ is assumed to be non-increasing, strictly positive with limit zero and $\mathfrak{a}(1)=1$. Thereby, we conclude that $\mathcal{R}^\star_n(\mathfrak{a},\Lambda) = \mathfrak{o}_n(1)$ and $n\mathcal{R}^\star_n(\mathfrak{a},\Lambda)\to\infty$ as well as $m^\star_n(\mathfrak{a})\in[\![1,n]\!]$ for all $n\in\mathbb{N}$. It follows also that $m^\star_n(\mathfrak{a}) = \arg\min\{\mathcal{R}^m_n(\mathfrak{a},\Lambda), m\in[\![1,n]\!]\}$ and $\mathcal{R}^\star_n(\mathfrak{a},\Lambda) = \min\{\mathcal{R}^m_n(\mathfrak{a},\Lambda), m\in[\![1,n]\!]\}$ for all $n\in\mathbb{N}$. We shall stress that in this situation the rate $\mathcal{R}^\star_n(\mathfrak{a},\Lambda)$ is non-parametric.* □

Let us now briefly illustrate the last definitions by stating the order of $m_n^\star(\mathfrak{a}, \Lambda)$ and $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda)$ for typical choices of the sequence $\mathfrak{a}$.

Numerical discussion 1.3.4.
We will illustrate all our results considering the following two configurations for the sequence $\mathfrak{a}$. Let

(o) $\mathfrak{a}(m)^2 \approx m^{-2p}$ with $p > 1$;

(s) $\mathfrak{a}(m)^2 \approx \exp(-m^{2p})$ with $p > 0$.

We consider as in Num. discussion 1.3.1 the situations **[o-o]**, **[o-s]** and **[s-o]**.

[o-o] $\mathcal{R}_n^{m_n^\star}(\mathfrak{a}, \Lambda) \approx (m_n^\star)^{-2p} \approx (m_n^\star)^{2a+1} n^{-1}$, and hence,

$m_n^\star(\mathfrak{a}) \approx n^{1/(2p+2a+1)}$ and $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) \approx n^{-2p/(2p+2a+1)}$

[o-s] $\mathcal{R}_n^{m_n^\star}(\mathfrak{a}, \Lambda) \approx (m_n^\star)^{-2p} \approx (m_n^\star)^{-(1-2a)_+} \exp((m_n^\star)^{2a}) n^{-1}$, and hence,
$m_n^\star(\mathfrak{a}) \approx (\log n - \frac{2p-(1-2a)_+}{2a} \log\log n)^{1/(2a)}$ and $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) \approx (\log n)^{-p/a}$.

[s-o] $\mathcal{R}_n^{m_n^\star}(\mathfrak{a}, \Lambda) \approx \exp(-(m_n^\star)^{2p}) \approx (m_n^\star)^{2a+1} n^{-1}$, and hence,
$m_n^\star(\mathfrak{a}) \approx (\log n - \frac{2a+1}{2p} \log\log n)^{1/(2p)}$ and $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) \approx (\log n)^{(2a+1)/(2p)} n^{-1}$. □

Unknown operator

Consider now the case where the operator is unknown. For all $n_\lambda \in \mathbb{N}$ we define

$$\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) := \max_{s \in \mathbb{N}} \{\mathfrak{a}(s)^2 [1 \wedge \Lambda(s)/n_\lambda]\}. \tag{1.6}$$

then for all $n_\lambda \in \mathbb{N}$ holds $\sup_{\theta^\circ \in \Theta(\mathfrak{a}, r)} \mathcal{R}_{n_\lambda}^\star(\theta^\circ, \Lambda) \leqslant r^2 \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda)$, since for all $\theta^\circ \in \Theta(\mathfrak{a}, r)$

$$\mathcal{R}_{n_\lambda}^\star(\theta^\circ, \Lambda) = \sum_{s \in \mathbb{N}} |\theta^\circ(s)|^2 [1 \wedge \Lambda(s)/n_\lambda] \leqslant \max_{s \in \mathbb{N}} \{\mathfrak{a}(s)^2 \min(1, \Lambda(s)/n_\lambda)\} \|\theta^\circ\|_{1/\mathfrak{a}}^2. \tag{1.7}$$

It follows for all $m, n, n_\lambda \in \mathbb{N}$ immediately that

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m}}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant (r^2 + 8)\mathcal{R}_n^m(\mathfrak{a}, \Lambda) + 8(\mathcal{C}_4 + 1) r^2 \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda). \tag{1.8}$$

The upper bound in the last display depends on the dimension parameter m and hence by choosing an optimal value $m_n^\star$ as in (1.5) the upper bound will be minimised, that is

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m_n^\star}}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant (r^2 + 8)\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) + 8(\mathcal{C}_4 + 1) r^2 \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda). \tag{1.9}$$

Numerical discussion 1.3.5.
Consider as in Num. discussion 1.3.4 the usual behaviours **[o-o]**, **[o-s]** and **[s-o]** for the sequences $(\mathfrak{a}(m))_{m \in \mathbb{N}}$ and $(\Lambda(m))_{m \in \mathbb{N}}$, where we have derived in Num. discussion 1.3.4 the corresponding minimax rates $(\mathcal{R}_n^\star(\mathfrak{a}, \Lambda))_{n \in \mathbb{N}}$, while for the rate $(\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda))_{n_\lambda \in \mathbb{N}}$ we get:

[o-o] $\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) \approx n_\lambda^{-(p \wedge a)/a}$

[o-s] $\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) \approx (\log n_\lambda)^{-p/a}$.

[s-o] $\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) \approx n_\lambda^{-1}$. □

Remark 1.3.5 *Since the operator T is not known, it is natural to consider a maximal risk also over a class for λ characterising the behaviour of $(\Lambda(s) = |\lambda(s)|^{-2})_{s\in\mathbb{N}}$, precisely $\mathcal{E}_{\mathfrak{e}}^{d} := \{\lambda \in l_2 : d^{-2} \leqslant \mathfrak{e}_s|\lambda|s||^2 = \mathfrak{e}_s/\Lambda(|s|) \leqslant d^2,\ \forall s \in \mathbb{N}\} \cap \mathcal{D}$. We shall note that for all $m \in \mathbb{N}$ and any $\lambda \in \mathcal{E}_{\mathfrak{e}}^{d}$, $d^{-2} \leqslant \Lambda_{+}(m)/\mathfrak{e}_{(m)} \leqslant d^2$, $d^{-2} \leqslant \Lambda_{\circ}(m)/\bar{\mathfrak{e}}_m \leqslant d^2$. Setting for all $n, n_\lambda \in \mathbb{N}$*

$$\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e}) := [\mathfrak{a}(m)^2 \vee m\bar{\mathfrak{e}}_m n^{-1}], \qquad m_n^\star(\mathfrak{a},\mathfrak{e}) := \arg\min\{\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e}), m \in \mathbb{N}\},$$
$$\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) := \mathcal{R}_n^{m_n^\star(\mathfrak{a},\mathfrak{e})}(\mathfrak{a},\mathfrak{e}) = \min\{\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e}), m \in \mathbb{N}\} \quad \text{and}$$
$$\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) := \max\{\mathfrak{a}(s)\min(1,\mathfrak{e}_s/n_\lambda), s \in \mathbb{N}\}. \tag{1.10}$$

we have

$$\mathcal{R}_n^\star(\mathfrak{a},\Lambda) = \min_{m\in\mathbb{N}}\{[\mathfrak{a}(m) \vee m\Lambda_\circ(m)n^{-1}]\} \leqslant d^2 \min_{m\in\mathbb{N}}\{[\mathfrak{a}(m) \vee m\bar{\mathfrak{e}}_m n^{-1}]\} \leqslant d^2\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e})$$
$$\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\Lambda) = \max_{s\in\mathbb{N}}\{\mathfrak{a}(s)^2[1 \wedge \Lambda(s)/n_\lambda]\} \leqslant d^2\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}). \tag{1.11}$$

It follows for all $m, n \in \mathbb{N}$ immediately that

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m}}, \Theta(\mathfrak{a},r), \mathcal{E}_{\mathfrak{e}}^{d}\big) \leqslant (r^2 + 8d^2)\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) + 8(\mathcal{C}_4+1)d^2r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\mathfrak{e}). \tag{1.12}$$

Johannes and Schwarz (2013a) have shown that $\inf_{\widehat{\theta}}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}, \Theta(\mathfrak{a},r), \mathcal{E}_{\mathfrak{e}}^{d}\big)$, where the infimum is taken over all possible estimators $\widehat{\theta}$ of θ°, is up to a constant bounded from below by $\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) \vee \mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\mathfrak{e})$. Consequently, the rate $(\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) \vee \mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\mathfrak{e}))_{n\in\mathbb{N}}$, the dimension parameters $(m_n^\star(\mathfrak{a}))_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m_n^\star(\mathfrak{a})}})_{n\in\mathbb{N}}$, respectively, is a minimax rate, a minimax dimension and minimax optimal (up to a constant). □

1.4 Bayesian approach

In the Bayesian paradigm, one does not assume the existence of a true parameter θ° but, defining a sigma algebra on Θ denoted $\mathcal{B}$, that this parameter is a random variable $\boldsymbol{\theta}$. Before observing any data, one might already have some knowledge or expectations about the said parameter and this knowledge is represented by the so-called prior distribution on $(\Xi, \mathcal{B})$, denoted $\mathbb{P}_{\boldsymbol{\theta}}$. Then, the data we observe, a random variable Y taking values in $(\mathbb{Y}, \mathcal{Y})$, depends on the parameter $\boldsymbol{\theta}$ in a way which is described by the conditional distribution $\mathbb{P}_{Y|\boldsymbol{\theta}}$ which we previously denoted $\mathbb{P}_\theta$ in the frequentist framework.
One would then be interested in the so-called posterior distribution which is, if it exists, the conditional distribution of the parameter $\boldsymbol{\theta}$ given that the data Y is y, that is to say, any function $\mathbb{P}_{\boldsymbol{\theta}|Y=y} : \mathcal{B} \times \mathbb{Y} \to [0,1]$ such that, for any A in $\mathcal{Y}$ $\mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y}(B)\mathbb{1}_{\{Y\in A\}}] = \mathbb{E}[\mathbb{1}_{\{\boldsymbol{\theta}\in B\}}\mathbb{1}_{\{Y\in A\}}]$. While in the parametric framework where Θ is a finite dimensional space, the existence of a satisfying posterior distribution (in a sense to be clarified later) is immediate, a deeper discussion is required in the non-parametric case. This topic is clearly treated in Ghosal and van der Vaart (2017) and we hence refer the reader to this book for more details. In this thesis, we will only consider models where Θ is a Polish space, and $\mathcal{B}$ is the associated Borel σ-algebra which ensures the existence of a satisfying

posterior distribution in the sense that it is a Markov Kernel, that is to say, for any Y in $\mathbb{Y}$, $B \mapsto \mathbb{P}_{\boldsymbol{\theta}|Y}(B)$ is a probability measure; and for any B in $\mathcal{B}$, the map $y \mapsto \mathbb{P}_{\boldsymbol{\theta}|Y=y}(B)$ is measurable.
Moreover, we will consider cases where the family of distributions $(\mathbb{P}_{Y|\boldsymbol{\theta}})_{\theta\in\Theta}$ is dominated by a common measure $\mathbb{P}^\circ$. We then define, for any θ in Θ, the Radon-Nikodym density of $\mathbb{P}_{Y|\theta}$ with respect to $\mathbb{P}^\circ$, as any function from $\mathbb{Y}$ to $[0,\infty[$, denoted $\mathrm{d}\mathbb{P}_{Y|\theta} / \mathrm{d}\mathbb{P}^\circ$, such that for any A in $\mathcal{Y}$, we have $\mathbb{P}_{Y|\theta}(A) = \int_A \mathrm{d}\mathbb{P}_{Y|\theta} / \mathrm{d}\mathbb{P}^\circ \mathrm{d}\mathbb{P}^\circ$. Hence, a regular version of the posterior distribution is given by the Bayes formula, that is, for any B in $\mathcal{B}$, $\mathbb{P}_{\boldsymbol{\theta}|Y}(B) = \frac{\int_B (\mathrm{d}\mathbb{P}_{Y|\boldsymbol{\theta}} / \mathrm{d}\mathbb{P}^\circ)(y) \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int (\mathrm{d}\mathbb{P}_{Y|\boldsymbol{\theta}} / \mathrm{d}\mathbb{P}^\circ)(y) \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta})}$.

1.4.1 Iteration procedure, self informative limit and Bayes carrier

As we have seen, the Bayesian paradigm relies on the notion of prior information. In some cases, it is reasonable to accept the eventuality that one does not trust the prior information available. A way to take in consideration such a possibility is the iteration procedure, studied for example in Bunke and Johannes (2005). It consists in considering the distribution of $\boldsymbol{\theta}|Y$ conditionally on Y. We then obtain, for any B in $\mathcal{B}$ the posterior distribution $\mathbb{P}_{\boldsymbol{\theta}|Y,Y}(B) = \frac{\int_B (\mathrm{d}\mathbb{P}_{Y|\boldsymbol{\theta}} / \mathrm{d}\mathbb{P}^\circ)(y) \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta})}{\int (\mathrm{d}\mathbb{P}_{Y|\boldsymbol{\theta}} / \mathrm{d}\mathbb{P}^\circ)(y) \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta})}$ which we may note $\mathbb{P}^{(2)}_{\boldsymbol{\theta}|Y}(B)$. Iterating this procedure generalises the *iterated posterior distribution*, given, for any η greater than 1 by $\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}|Y}(B) = \frac{\int_B (\mathrm{d}\mathbb{P}_{Y|\boldsymbol{\theta}} / \mathrm{d}\mathbb{P}^\circ)(y) \mathrm{d}\mathbb{P}^{(\eta-1)}_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta})}{\int (\mathrm{d}\mathbb{P}_{Y|\boldsymbol{\theta}} / \mathrm{d}\mathbb{P}^\circ)(y) \mathrm{d}\mathbb{P}^{(\eta-1)}_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta})}$.
Then, a question of interest is the asymptotic with respect to the iteration parameter η. The support of the limiting distribution (that is to say, the smallest closed set with probability 1), is called self informative Bayes carrier, whereas that posterior mean of this limiting distribution is called self informative limit.

1.4.2 Typical priors for non-parametric models

In the context we depicted previously, many priors have been considered. We will however focus on a family of priors named Gaussian sieve priors and an *adaptive* variant of theirs. They are a specific case of the Gaussian process prior family, defined hereafter , and for which an in depth and practical presentation can be found in Rasmussen (2003).

Definition Given a sequence $\theta^\times$ in Θ and a semi-definite positive operator on Ξ^2 Ξ, let be $\boldsymbol{f}$ a $(\Xi, \mathcal{B})$-valued random variable and $\mathbb{P}_{\boldsymbol{f}}$ its distribution. If $\mathbb{P}_{\boldsymbol{f}}$ is such that, for any integer p and any collection $(x_j)_{j\in[\![1,p]\!]}$ in Ξ^p, the collection of random variables $(\langle \boldsymbol{f} | x_j\rangle_\Xi)_{j\in[\![1,p]\!]}$ is a Gaussian vector with mean $(\sum_{s\in\mathbb{F}} \theta^\times(s)\overline{[x_j]}(s) e_s)_{j\in[\![1,p]\!]}$, and covariance matrix $(\Sigma(x_j, x_l))_{(j,l)\in[\![1,p]\!]}$, we say that $\mathbb{P}_{\boldsymbol{f}}$ is a Gaussian process prior.
We will in particular be interested in the distribution $\mathbb{P}_{\boldsymbol{\theta}}$, the distribution induced on Θ by $\mathbb{P}_{\boldsymbol{f}}$. For any collection $(s_j)_{j\in[\![1,p]\!]}$ in $\mathbb{F}^p$, the vector of random variables $(\boldsymbol{\theta}(s_j))_{j\in[\![1,p]\!]} = (\langle \boldsymbol{f} | e_{s_j}\rangle_\Xi)_{j\in[\![1,p]\!]}$ follows a normal distribution with mean $(\theta^\times(s_j))_{j\in[\![1,p]\!]}$, and covariance matrix $(\Sigma(e_{s_j}, e_{s_k})))_{(j,k)\in[\![1,p]\!]^2}$. □

In this case, the prior distribution is entirely determined by the choice of $\theta^\times$ and Σ. We will in practice only consider cases when $\mathbb{F} = \mathbb{N}$ and hence, we can chose $\Sigma(e_s, e_{s'})$ to be 0 as soon as $s \neq s'$.

Definition 27 The Gaussian sieve priors are Gaussian process priors for which there exists m in $\mathbb{M}$ such that for any s in $\mathbb{F}$, $\Sigma(e_s, e_s) = \mathbb{1}_{\{s \in \mathbb{F}_m\}}$ where $(\mathbb{F}_m)_{m \in \mathbb{M}}$ is a sieve. $\square$

Here we will have $(\mathbb{F}_m)_{m \in \mathbb{M}} = ([\![-m, m]\!])_{m \in \mathbb{N}}$ or $(\mathbb{F}_m)_{m \in \mathbb{M}} = ([\![0, m]\!])_{m \in \mathbb{N}}$.

1.4.3 The pragmatic Bayesian approach

Even though the traditional Bayesian approach does not admit the existence of a true parameter and focuses on the study of the posterior distribution, one could wonder about the performances of such methods under a frequentist lens. Admitting the existence of a true parameter θ° and assuming prior knowledge $\mathbb{P}_{\boldsymbol{\theta}}$ about it and observing some data Y which distribution is hence given by $\mathbb{P}_{Y|\boldsymbol{\theta}=\theta^\circ}$, we wonder if the posterior distribution contracts around θ°, as formulated in the following definition.

Definition 28 A posterior distribution is said to be consistent at θ° if, for any real, strictly positive, constant c we have $\lim_{n\to\infty} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\theta^\circ - \boldsymbol{\theta}\|_{l^2}^2 \geqslant c)] = 0$. $\square$

Notice that for any c, the probability $\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\theta^\circ - \boldsymbol{\theta}\|_{l^2}^2 \geqslant c)$ is a random variable which depends on the observations Y^n. In addition, if, for any Y in $\mathbb{Y}$, the measure $\mathbb{P}_{\boldsymbol{\theta}|Y}$ is a Dirac measure, say $\delta_{\theta(Y)}$ then we recover the definition of the probability for the estimator $\theta(Y)$ to exceed the loss c, indeed,

$$\mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\theta^\circ - \boldsymbol{\theta}\|_{l^2}^2 \geqslant c)] = \mathbb{E}[\mathbb{1}_{(\|\theta^\circ - \theta(Y)\|_{l^2}^2 \geqslant c)}] = \mathbb{P}(\|\theta^\circ - \theta(Y)\|_{l^2}^2 \geqslant c).$$

One can then quantify the so called rate of contraction of the posterior distribution.

Definition 29 Given a consistent posterior distribution $\mathbb{P}_{\boldsymbol{\theta}|Y^n}$, a sequence $(\Psi_n)_{n\in\mathbb{N}}$ of real, strictly positive, numbers, converging to 0 is called a contraction rate for $\mathbb{P}_{\boldsymbol{\theta}|Y^n}$ if for any increasing unbounded sequence $(c_n)_{n\in\mathbb{N}}$, $\lim_{n\to\infty} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\theta^\circ - \boldsymbol{\theta}\|_{l^2}^2 \geqslant c_n\Psi_n)] = 0$.
If, in addition, for any increasing unbounded sequence $(c_n)_{n\in\mathbb{N}}$ we also have
$\lim_{n\to\infty} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\theta^\circ - \boldsymbol{\theta}\|_{l^2}^2 \leqslant c_n^{-1}\Psi_n)] = 0$ then, $(\Psi_n)_{n\in\mathbb{N}}$ is called an exact contraction rate for $\mathbb{P}_{\boldsymbol{\theta}|Y^n}$. $\square$

Hence, considering a family of posterior distributions $\mathcal{G}$, such as the one obtained while using the sieve priors introduced earlier, we can define the notion of oracle optimal prior within this family at a certain true parameter θ°.

Definition 30 A posterior distribution $\mathbb{P}_{\boldsymbol{\theta}|Y^n}$ belonging to a family $\mathcal{G}$ of posterior distributions is called oracle optimal at a true parameter value θ° if there exists a sequence $\Phi_n^\circ(\theta^\circ)$ such that, for any increasing and unbounded sequence $(c_n)_{n\in\mathbb{N}}$, we have $\sup_{\mathbb{Q}\in\mathcal{G}} \lim_{n\to\infty} \mathbb{E}[\mathbb{Q}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2 \leqslant c_n^{-1}\Psi_n^\circ(\theta^\circ))] = 0$ and in addition
$\lim_{n\to\infty} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2 \leqslant c_n\Psi_n^\circ(\theta^\circ))] = 1$. $\square$

As in the frequentist case, one can alternatively consider uniform rates.

Definition 31 Considering a subspace of the parameter space $\widetilde{\Theta}$, and a posterior distribution $\mathbb{P}_{\boldsymbol{\theta}|Y}$, a sequence $(\Psi_n^\star(\widetilde{\Theta}))_{n\in\mathbb{N}}$ is called uniform contraction rate is, for any increasing unbounded sequence $(c_n)_n$ we have $\lim_{n\to\infty} \inf_{\theta^\circ\in\widetilde{\Theta}} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2 \leqslant c_n\Psi_n^\star(\widetilde{\Theta}))] = 1$.
It is called an exact uniform contraction rate if in addition, for any increasing unbounded sequence, $\lim_{n\to\infty} \inf_{\theta^\circ\in\widetilde{\Theta}} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2 \leqslant c_n^{-1}\Psi_n^\star(\widetilde{\Theta}))] = 0$. $\square$

And with this definition at hand, we can define the minimax optimality, in a similar manner to the frequentist notion.

Definition 32 A sequence $(\Psi^{\star}_n(\widetilde{\Theta}))_{n\in\mathbb{N}}$ is called minimax optimal contraction rate if, for any increasing unbounded sequence $(c_n)_{n\in\mathbb{N}}$, we have

$$\lim_{n\to\infty} \sup_{\mathbb{Q}_{\boldsymbol{\theta}|Y^n}} \inf_{\theta^\circ\in\widetilde{\Theta}} \mathbb{E}[\mathbb{Q}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta}-\theta^\circ\|^2_{l^2} \leqslant c_n^{-1}\Psi^{\star}_n(\widetilde{\Theta}))]$$

and at the same time there exists a posterior distribution $\mathbb{P}_{\boldsymbol{\theta}|Y^n}$ such that for any increasing unbounded sequence $(c_n)_{n\in\mathbb{N}}$ we have

$$\lim_{n\to\infty} \inf_{\theta^\circ\in\widetilde{\Theta}} \mathbb{E}[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta}-\theta^\circ\|^2_{l^2} \leqslant c_n\Psi^{\star}_n(\widetilde{\Theta}))];$$

such a posterior distribution is called minimax optimal. Note that in "$\sup_{\mathbb{Q}_{\boldsymbol{\theta}|Y^n}}$", the supremum is taken over any possible posterior distribution contrarily to the definition of the oracle optimality. □

1.4.4 Existing central results

Consistence and contraction rate of posterior distributions have gathered interest for a long while one could mention the work of Doob (1949). More recently, results introduced in Schwartz (1965) and reformulated in Ghosal et al. (2000), seems to be at the origin of a very fecund theory allowing the systematic study of the posterior contraction rates based on the complexity of the parameter space in terms of ε-packing numbers. Originally formulated with Kullback-Leibler divergence, these results have been adapted, for example, to l^p distances in Giné and Nickl (2011). It has since been applied to many models, including inverse problems (Knapik et al. (2011)).

One of the main limitations of this approach is that the contraction rates obtained are generally penalised by a log-loss compared to the convergence rates of frequentist approaches. That is why, in this thesis, we will give more attention to the approach suggested in Johannes et al. (2014) which allowed, in the context of the inverse Gaussian sequence space model, to obtained exact contraction rate for the posterior distribution obtained with a Gaussian sieve prior.

1.5 Inverse Gaussian sequence space model

Let Ξ be space of function from $[0,1]$ to $\mathbb{C}$. We equip the space with internal addition $+$ such that for any x and y in Ξ, $x+y=(t\mapsto x(t)+y(t))$, external product $\cdot$ such that for any a in $\mathbb{C}$, $a\cdot x=(t\mapsto a\cdot x(t))$ and inner product $\langle x|y\rangle_{L^2}=\int x(t)\cdot\overline{y}(t)\,\mathrm{d}t$. Hence, Ξ equipped with the L^2-norm generated by $\langle\cdot|\cdot\rangle_{L^2}$ is a Hilbert space, for which the family of functions $(e_s)_{s\in\mathbb{Z}}$ such that, for any s in $\mathbb{Z}$, $e_s: t\mapsto\exp[-2\imath\pi st]$ is an orthonormal basis. Then, denote $\mathbb{L}^2$ the sub-space of Ξ of square integrable, **real-valued** functions defined on $[0,1]$. We define on $\mathbb{L}^2$ the convolution product $\star$ such that for any x and y in $\mathbb{L}^2$,

we have $x \star y : t \mapsto \int_{[0,1]} x(t - s - \lfloor t - s \rfloor) y(s) \, \mathrm{d}s$. Let also be Θ, the space of $\mathbb{Z}$-indexed, $\mathbb{C}$-valued sequences, equipped with internal addition $+$ such that for any $[x]$ and $[y]$ in Θ, $[x] + [y] = (s \mapsto [x](s) + [y](s))$; external product $\cdot$ such that, for any a in $\mathbb{C}$, $a \cdot [x] = (s \mapsto a \cdot [x](s))$; and inner product $\langle [x] | [y] \rangle_{l^2} = \sum_{s \in \mathbb{Z}} [x](s) \overline{[y]}(s)$. Hence, Θ equipped with the l^2-norm derived from $\langle \cdot | \cdot \rangle_{l^2}$ is a Hilbert space, for which the family $(s \mapsto \mathbb{1}_{\{s = s^\star\}})_{s^\star \in \mathbb{Z}}$ is an orthonormal basis.
In this context, we have $\mathcal{F}$, the Fourier transform with respect to $(e_s)_{s \in \mathbb{Z}}$, that is to say, $\mathcal{F} : \Xi \to \Theta, \quad x \mapsto [x] (= s \mapsto \langle x | e_s \rangle_{L^2})$. Notice that, for any element x of $\mathbb{L}^2$, due to the fact that it is real valued, we have for any s in $\mathbb{Z}$, that $[x](s) = \overline{[x]}(-s)$, and due to the fact that it is square integrable, $[x]$ is square summable and we hence denote $\mathcal{L}^2$ the subspace of Θ of square summable sequences $[x]$ such that, for any s in $\mathbb{Z}$, $[x](s) = \overline{[x]}(-s)$. Also, for any x and y in $\mathbb{L}^2$, we have $[x \star y] = [x] \cdot [y]$. Due to the fact that the Fourier transform is unitary, we also have, for any t in $[0,1]$ and x in $\mathbb{L}^2$ that $x(t) = \mathcal{F}^\star([x]) = \sum_{s \in \mathbb{Z}} [x](s) e_s(t)$.

Keeping in mind the notations used until here, let f and h be in $\mathbb{L}^2$ and $g := f \star h$, hence, we have $T : \mathbb{L}^2 \to \mathbb{L}^2, \quad x \mapsto x \star h$ and we have three elements of $\mathcal{L}^2$ given by, $\theta^\circ = \mathcal{F}(f)$, $\lambda = \mathcal{F}(h)$, and $\phi = \mathcal{F}(g) = \theta^\circ \cdot \lambda$. Remind that for any x in $\mathbb{L}^2$, we have $\|x\|_{L^2} = \|[x]\|_{l^2}$ and hence we will study estimation procedures for θ°, however, the reader should keep in mind that it is motivated by the estimation of f, which has equivalent performances due to Plancherel theorem.

1.5.1 Known operator

Let Y be the Gaussian process such that for any t in $[0,1]$, we have $\mathrm{d}Y(t) = \mathrm{d}W(t) + g(t) \, \mathrm{d}t$ where W is the Brownian motion. Hence, there exist sequences of real-valued random variables $(\xi_1(s))_{s \in \mathbb{Z}}$ and $(\xi_2(s))_{s \in \mathbb{Z}}$ such that for any s and s' in $\mathbb{Z}$, we have $\int_{[0,1]} e_s(t) \, \mathrm{d}Y(t) = \int_{[0,1]} \cos(2\pi s t) \, \mathrm{d}W(t) + \imath \int_{[0,1]} \sin(2\pi s t) \, \mathrm{d}W(t) + \phi(s) = \xi_1(s) + \imath \xi_2(s) + \phi(s)$ with $\xi_1(s) \sim \mathcal{N}(0, 1/2)$, $\xi_2(s) \sim \mathcal{N}(0, 1/2)$, and $\xi_1(s)$ is independent of $\xi_2(s)$, in addition, $\mathbb{C}\mathrm{ov}(\xi_1(s), \xi_1(s')) = \mathbb{1}_{\{|s'| = |s|\}}$ and $\mathbb{C}\mathrm{ov}(\xi_2(s), \xi_2(s')) = \mathrm{Sign}(s \cdot s') \mathbb{1}_{\{|s'| = |s|\}}$.

Define the iid. stochastic process $(Y_p)_{p \in \mathbb{Z}}$ such that, for any p in $\mathbb{Z}$, Y_p is identically distributed to Y. We observe the sub-vector $Y^n = (Y_p)_{p \in [\![1,n]\!]}$ of $(Y_p)_{p \in \mathbb{Z}}$ and define, for any s in $\mathbb{N}$ the estimates $\phi_n(s) = \sum_{p=1}^n \int_{[0,1]} e_s(t) \, \mathrm{d}Y(t)/n$ which verifies $\mathfrak{Re}(\phi_n(s)) \sim \mathcal{N}(\mathfrak{Re}(\phi(s)), 2/n)$ and $\mathfrak{Im}(\phi_n(s)) \sim \mathcal{N}(\mathfrak{Im}(\phi(s)), 2/n)$ and as we assume Assumption 3, we know λ and for any s in $\mathbb{Z}$ $|\lambda(s)| > 0$ so we can define $\theta_n(s) := \phi_n(s)/\lambda(s)$, which verifies $\mathfrak{Re}(\theta_n(s)) \sim \mathcal{N}(\mathfrak{Re}(\theta^\circ), \Lambda(s)/n)$; $\mathfrak{Im}(\theta_n(s)) \sim \mathcal{N}(\mathfrak{Im}(\theta^\circ), \Lambda(s)/n)$; and $\mathbb{C}\mathrm{ov}(\mathfrak{Re}(\theta_n(s)), \mathfrak{Im}(\theta_n(s))) = 0$. Finally, we define, for any m in $\mathbb{N}$, the projections estimator $\theta_{n,\overline{m}} = (\theta_n(s) \mathbb{1}_{\{|s| \leqslant m\}})_{s \in \mathbb{Z}}$. We give, in fig. 1.1 an illustration of a projection estimator.
Notice that, for any m in $\mathbb{N}$, $\mathbb{E}[\|\theta^\circ - \theta_{n,\overline{m}}\|_{l^2}^2] = 2 \sum_{s \in \mathbb{N}} \mathbb{E}[(\mathfrak{Re}(\theta^\circ(s)) - \mathfrak{Re}(\theta_{n,\overline{m}}(s)))^2] + 2 \sum_{s \in \mathbb{N}} \mathbb{E}[(\mathfrak{Im}(\theta^\circ(s)) - \mathfrak{Im}(\theta_{n,\overline{m}}(s)))^2]$. Hence, real and imaginary parts can be treated separately in an identical manner and considering the positive indexes is sufficient and we only will give attention to the estimation of the real part of the positively indexed coefficients of θ° and we give this final formulation for the model.

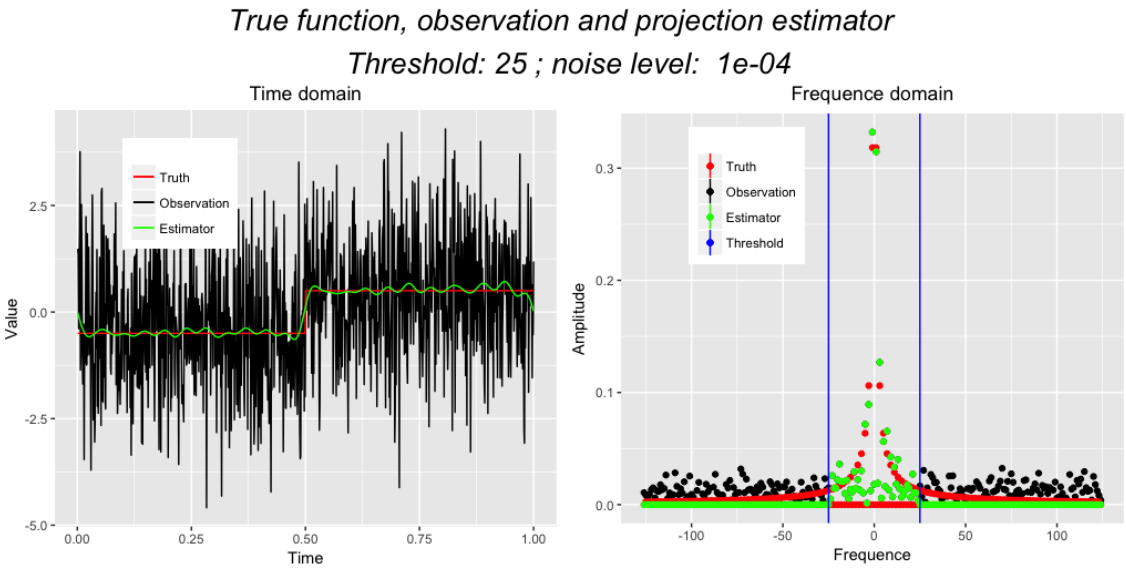

Figure 1.1: Projection estimator in the time and frequency space, direct problem case.

DEFINITION 33 Let Θ be the space of $\mathbb{N}^{\star}$-indexed, $\mathbb{R}$-valued sequences, equipped with the inner product $\langle \cdot | \cdot \rangle_{l^2} : ([x],[y]) \mapsto \sum_{s\in\mathbb{N}} [x](s) \cdot [y](s)$ and the associated norm $\| \cdot \|_{l^2} : [x] \mapsto \sum [x](s)^2$. Let $\mathcal{L}^2$ be the subspace of Θ of square-summable sequences.
Given three elements of Θ denoted θ°, λ, and ϕ, such that, $\phi = \theta^{\circ} \cdot \lambda$; for any s in $\mathbb{N}$, $0 < |\lambda(s)| \leqslant 1$; and θ° is an element of $\mathcal{L}^2$.
We observe ϕ_n in Θ such that, for any s and s' in $\mathbb{N}$ such that $s \neq s'$, we have $\phi_n(s) \sim \mathcal{N}(\phi(s), n^{-1})$ and $\mathbb{C}\mathrm{ov}(\phi_n(s), \phi_n(s')) = 0$. □

The likelihood for this model is given by

$$L(\phi_n, \theta) \propto \exp[n^{-1}(\sum_{s\in\mathbb{N}} \phi_n(s)\theta(s)\lambda(s) - \sum_{s\in\mathbb{N}} \Lambda(s)^{-1}\theta(s)^2/2)].$$

Notice that as in Assumption 9, $\mathbb{V}[\langle Y_0 | e_s \rangle_{L^2}] = 1$, hence, all the results obtained considering the convergence rates remain true. Hence let us give the following reminders.

NOTATION For any m in $\mathbb{N}^{\star}$; s in $\mathbb{N}^{\star}$; and θ in Θ, let be the following quantities:

$$\mathfrak{b}_m^2(\theta) := \|\theta_{\underline{0}}\|_{l^2}^{-2} \|\theta_{\underline{m}}\|^2; \quad \Lambda(s) = |\lambda^{-1}(s)|^2;$$
$$\Lambda_{\circ}(m) = m^{-1} \sum_{0<s\leqslant m} \Lambda(s); \quad \Lambda_{+}(m) := \max_{s\in[\![1,m]\!]} \{\Lambda(s)\}.$$

We then denote in the following way the risk for projection estimators:

$$\mathcal{R}_n^m(\theta^{\circ}, \Lambda) := [n^{-1} m \Lambda_{\circ}(m) \vee \mathfrak{b}_m^2(\theta^{\circ})].$$

Which gives us the following oracle rates,

$$m_n^{\circ} \in \arg\min_{m\subset\mathbb{M}} \{\mathcal{R}_n^m(\theta, \lambda)\} = \arg\min_{m\subset\mathbb{M}} \left\{[n^{-1} m \Lambda_{\circ}(m) \vee \mathfrak{b}_m^2(\theta^{\circ})]\right\};$$
$$\mathcal{R}_n^{\circ}(\theta^{\circ}, \Lambda) = \mathcal{R}_n^{m_n^{\circ}}(\theta^{\circ}, \Lambda) = \min_{m\in\mathbb{N}} \mathcal{R}_n^m(\theta^{\circ}, \Lambda).$$

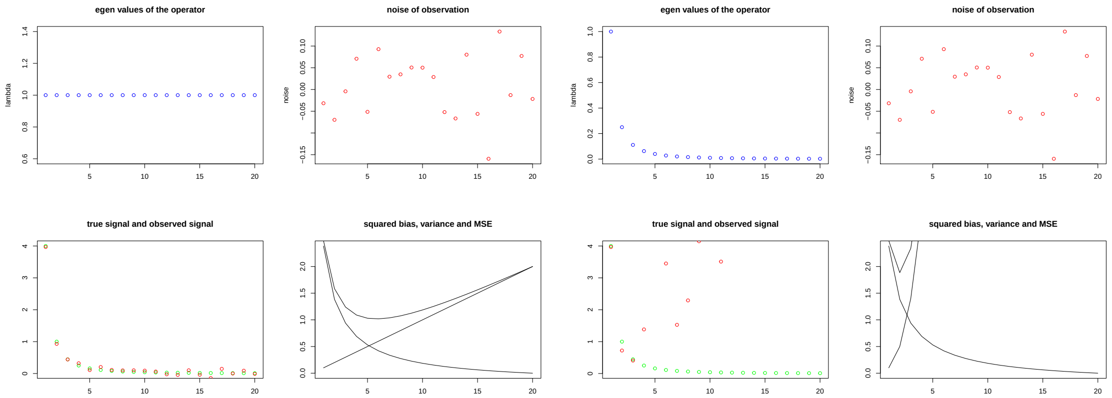

Figure 1.2: Influence of the operator Eigen-values sequence decay on the quadratic risk

And we were able to obtain the following maximal rates.

$$\mathcal{R}_n^m(\mathfrak{a},\Lambda) := [\mathfrak{a}(m)^2 \vee m\Lambda_\circ(m)n^{-1}] := \max\left(\mathfrak{a}(m)^2, m\Lambda_\circ(m)n^{-1}\right),$$
$$m_n^\star(\mathfrak{a}) := m_n^\star(\mathfrak{a},\Lambda) := \arg\min\{\mathcal{R}_n^m(\mathfrak{a},\Lambda), m\in\mathbb{N}\} \quad \text{and}$$
$$\mathcal{R}_n^\star(\mathfrak{a},\Lambda) := \mathcal{R}_n^{m_n^\star(\mathfrak{a})}(\mathfrak{a},\Lambda) = \min\{\mathcal{R}_n^m(\mathfrak{a},\Lambda), m\in\mathbb{N}\}. \quad (1.13)$$

□

Reminder *We shall emphasise that $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \geqslant n^{-1}$ for all $n\in\mathbb{N}$, and $\lim_{n\to\infty}\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = 0$. Observe that for all $\delta > 0$ there exists $m_\delta \in \mathbb{N}$ and $n_\delta \in \mathbb{N}$ such that for all $n \geqslant n_\delta$ holds $\mathfrak{b}_{m_\delta}^2(\theta^\circ) \leqslant \delta$ and $m_\delta\Lambda_\circ(m_\delta)n^{-1} \leqslant \delta$, and whence $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \leqslant \mathcal{R}_n^{m_\delta}(\theta^\circ,\Lambda) \leqslant \delta$. Moreover, we have $m_n^\circ \in [\![1,n]\!]$. Indeed, by construction holds $\mathfrak{b}_n^2(\theta^\circ) \leqslant 1 < (n+1)n^{-1} \leqslant (n+1)\Lambda_\circ(n+1)n^{-1}$, and hence $\mathcal{R}_n^n(\theta^\circ,\Lambda) < \mathcal{R}_n^m(\theta^\circ,\Lambda)$ for all $m \in [\![n+1,\infty[\![$ which in turn implies the claim $m_n^\circ \in [\![1,n]\!]$. Obviously, it follows thus $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \min\{\mathcal{R}_n^m(\theta^\circ,\Lambda), m\in[\![1,n]\!]\}$ for all $n \in \mathbb{N}$. We shall use those elementary findings in the sequel without further reference. The sequence $\mathcal{R}_n^\circ(\theta,\lambda)$ is then an exact oracle convergence rate and the projection estimator $\theta_{n,\overline{m_n^\circ}}$ is an oracle optimal estimator.*

*Note that, in case **(p)**, the oracle rate is parametric, that is $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \approx n^{-1}$. More precisely, if $\theta^\circ = 0$ then for each $m\in\mathbb{N}$, $\mathbb{E}\|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2 = m\Lambda_\circ(m)n^{-1}$, and hence $m_n^\circ = 1$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = 2\Lambda_\circ(1)n^{-1} \sim n^{-1}$. Otherwise if there is $K\in\mathbb{N}$ with $\mathfrak{b}_{K-1}(\theta^\circ) > 0$ and $\mathfrak{b}_K(\theta^\circ) = 0$, then setting $n_{\theta^\circ} := \frac{K\Lambda_\circ(K)}{\mathfrak{b}_{K-1}^2(\theta^\circ)}$, for all $n \geqslant n_{\theta^\circ}$ holds $\mathfrak{b}_{K-1}^2(\theta^\circ) > K\Lambda_\circ(K)n^{-1}$, and hence $m_n^\circ = K$ and $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = K\Lambda_\circ(K)n^{-1} \sim n^{-1}$. On the other hand side, in case **(np)** the oracle rate is non-parametric, more precisely, it holds $\lim_{n\to\infty} n\mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \infty$. Indeed, since $\mathfrak{b}_{m_n^\circ}^2(\theta^\circ) \leqslant \mathcal{R}_n^\circ(\theta^\circ,\Lambda) = \mathcal{R}_n^\circ(\theta^\circ,\Lambda) \in \mathfrak{o}_n(1)$ follows $m_n^\circ \to \infty$ and hence $m_n^\circ\Lambda_\circ(m_n^\circ) \to \infty$ which implies the claim because $n\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \geqslant m_n^\circ\Lambda_\circ(m_n^\circ)$.*

When considering the maximal rate, by construction it holds $\mathcal{R}_n^\star(\mathfrak{a},\Lambda) \geqslant n^{-1}$ for all $n\in\mathbb{N}$. The following statements can be shown using the same arguments as in Remark 1.3.1 by exploiting that the sequence $\mathfrak{a}$ is assumed to be non-increasing, strictly positive with limit zero

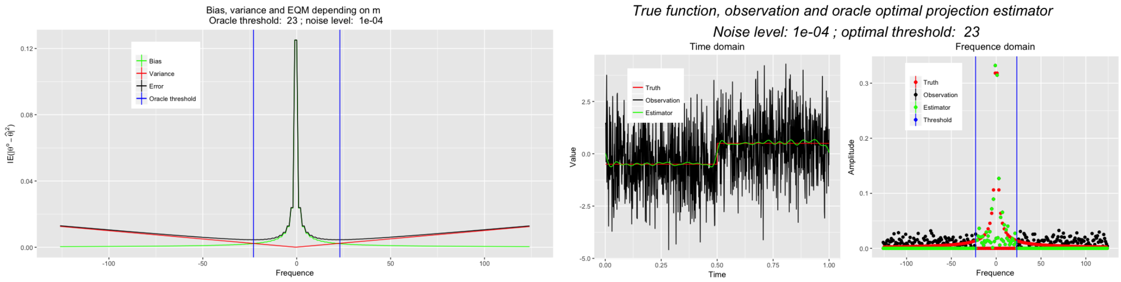

Figure 1.3: Illustration of the oracle value for m and associated projection estimator.

and $\mathfrak{a}(1) = 1$. *Thereby, we conclude that* $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) = \mathfrak{o}_n(1)$ *and* $n\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) \to \infty$ *as well as* $m_n^\star(\mathfrak{a}) \in [\![1, n]\!]$ *for all* $n \in \mathbb{N}$. *It follows also that* $m_n^\star(\mathfrak{a}) = \arg\min\{\mathcal{R}_n^m(\mathfrak{a}, \Lambda), m \in [\![1, n]\!]\}$ *and* $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) = \min\{\mathcal{R}_n^m(\mathfrak{a}, \Lambda), m \in [\![1, n]\!]\}$ *for all* $n \in \mathbb{N}$. *We shall stress that in this situation the rate* $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda)$ *is non-parametric.* □

We give in fig. 1.3 an illustration of the oracle estimator in a Gaussian sequence space model, in the direct problem case that is to say $\lambda(s) = 1$ for all s in $\mathbb{N}$.

1.5.2 Unknown operator

In the case of an unknown operator, that is, h, and hence, λ unknown, we supposed that, in addition to Y we define another Gaussian process ε. Let ε be the Gaussian process such that for any t in $[0,1]$, we have $\mathrm{d}\varepsilon(t) = \mathrm{d}W(t) + h(t)\,\mathrm{d}t$ where W is the Brownian motion. Hence, there exist sequences of real-valued random variables $(\xi_3(s))_{s\in\mathbb{Z}}$ and $(\xi_4(s))_{s\in\mathbb{Z}}$ such that for any s and s' in $\mathbb{Z}$, we have $\int_{[0,1]} e_s(t)\,\mathrm{d}\varepsilon(t) = \int_{[0,1]} \cos(2\pi st)\,\mathrm{d}W(t) + \imath \int_{[0,1]} \sin(2\pi st)\,\mathrm{d}W(t) + \lambda(s) = \xi_3(s) + \imath\xi_4(s) + \phi(s)$ with $\xi_3(s) \sim \mathcal{N}(0, 1/2)$, $\xi_4(s) \sim \mathcal{N}(0, 1/2)$, and $\xi_3(s)$ is independent of $\xi_4(s)$, in addition, $\mathbb{C}\mathrm{ov}(\xi_3(s), \xi_3(s')) = \mathbb{1}_{\{|s'|=|s|\}}$ and $\mathbb{C}\mathrm{ov}(\xi_4(s), \xi_4(s')) = \mathrm{Sign}(s \cdot s')\mathbb{1}_{\{|s'|=|s|\}}$. Define the iid. stochastic process $(\varepsilon_p)_{p\in\mathbb{Z}}$ such that, for any p in $\mathbb{Z}$, ε_p is identically distributed to ε. We observe the sub-vector $\varepsilon^{n_\lambda} = (\varepsilon_p)_{p\in[\![1,n_\lambda]\!]}$ of $(\varepsilon_p)_{p\in\mathbb{Z}}$ and define, for any s in $\mathbb{N}$ the estimates
$\lambda_{n_\lambda}(s) = \sum_{p=1}^{n_\lambda} \int_{[0,1]} e_s(t)\,\mathrm{d}\varepsilon(t) n_\lambda^{-1}$ which verifies $\mathfrak{Re}(\lambda_{n_\lambda}(s)) \sim \mathcal{N}(\mathfrak{Re}(\lambda(s)), 2n_\lambda^{-1})$ and $\mathfrak{Im}(\lambda_{n_\lambda}(s)) \sim \mathcal{N}(\mathfrak{Im}(\lambda(s)), 2n_\lambda^{-1})$ so we define $\theta_{n,n_\lambda}(s) := \phi_n(s)\lambda_{n_\lambda}^{+}(s)$, with $\lambda_{n_\lambda}^{+}(s) = \lambda_{n_\lambda}^{-1}(s)\mathbb{1}_{\{|\lambda_{n_\lambda}|^2 \geqslant n_\lambda^{-1}\}}$.
We may carry the same observations about the convergence rate and hence we give the following final definition for the model in the case of unknown operator.

Definition 34 Let Θ be the space of $\mathbb{N}$-indexed, $\mathbb{R}$-valued sequences, equipped with the inner product $\langle\cdot|\cdot\rangle_{l^2} : ([x], [y]) \mapsto \sum_{s\in\mathbb{N}}[x](s) \cdot [y](s)$ and the associated norm $\|\cdot\|_{l^2} : [x] \mapsto \sum[x](s)^2$. Let $\mathcal{L}^2$ be the subspace of Θ of square-summable sequences.
Given three elements of Θ denoted θ°, λ, and ϕ, such that, $\phi = \theta^\circ \cdot \lambda$; for any s in $\mathbb{N}$, $0 < |\lambda(s)| \leqslant 1$; and θ° is an element of $\mathcal{L}^2$.
We observe the Θ-valued random variable ϕ_n such that, for any s and s' in $\mathbb{N}$ such that $s \neq s'$, we have $\phi_n(s) \sim \mathcal{N}(\phi(s), n^{-1})$ and $\mathbb{C}\mathrm{ov}(\phi_n(s), \phi_n(s')) = 0$. On the other hand, we also observe the Θ-valued random variable λ_{n_λ} such that, for any s and s' in $\mathbb{N}$ such that $s \neq s'$, we have $\lambda_{n_\lambda}(s) \sim \mathcal{N}(\lambda(s), n_\lambda^{-1})$ and $\mathbb{C}\mathrm{ov}(\lambda_{n_\lambda}(s), \lambda_{n_\lambda}(s')) = 0$. □

As a direct consequence we have for any s in $\mathbb{N}^\star$, as in Assumption 9, $\lambda(s) - \lambda_{n_\lambda}(s) \sim \mathcal{N}(0, n_\lambda^{-1})$ and hence, $n_\lambda^2 \mathbb{E}|\lambda(s) - \lambda_{n_\lambda}(s)|^4 = 3$ and $n_\lambda \mathbb{E}[|\lambda_{n_\lambda}(s) - \lambda(s)|^2] = 1$. Hence, Assumption 9 holds true and the analysis of the optimal rates carried out previously still holds true. We hence remind the following results we obtained.

Notation In addition to the case of a known operator, we defined the following rate, for all n_λ in N

$$\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ, \Lambda) := \sum_{s\in\mathbb{F}} |\theta(s)|^2 (1 \wedge n_\lambda^{-1}\Lambda(s))$$

$$\mathcal{R}^\star_{n_\lambda}(\mathfrak{a}, \Lambda) := \max_{s\in\mathbb{N}} \{\mathfrak{a}(s)^2 [1 \wedge \Lambda(s)/n_\lambda]\} \|\theta^\circ\|^2_{1/\mathfrak{a}}.$$

Reminder *We have*

$$\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m^\circ_n}}) \leqslant (1 + \|\theta^\circ_{\underline{0}}\|^2_{l^2}) \mathcal{R}^\circ_n(\theta^\circ, \Lambda) + 2\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ, \Lambda).$$

We note that $\|\theta^\circ_{\underline{0}}\|^2_{l^2} = 0$ implies $\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ, \Lambda) = 0$, while for $\|\theta^\circ_{\underline{0}}\|^2_{l^2} > 0$ holds $\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ, \Lambda) \geqslant \sum_{s:\Lambda(s)>n_\lambda} |\theta^\circ(s)|^2 + n_\lambda^{-1} \sum_{s:\Lambda(s)\leqslant n_\lambda} |\theta^\circ(s)|^2 \geqslant n_\lambda^{-1} \sum_{s\in\mathbb{N}} |\theta^\circ(s)|^2 = \|\theta^\circ_{\underline{0}}\|^2_{l^2} n_\lambda^{-1}$, thereby whenever $\theta^\circ \neq 0$ any additional term of order $n^{-1} + n_\lambda^{-1}$ is negligible with respect to the rate $\mathcal{R}^\circ_n(\theta^\circ, \Lambda) + \mathcal{R}^\dagger_{n_\lambda}(\theta^\circ, \Lambda)$, since $\mathcal{R}^\circ_n(\theta^\circ, \Lambda) \geqslant n^{-1}$, which we will use below without further reference. We shall emphasise that in case $n = n_\lambda$ it holds

$$\begin{aligned}\mathcal{R}^\dagger_n(\theta^\circ, \Lambda) &= \sum_{0<s\leqslant m^\circ_n} |\theta^\circ(s)|^2 [1 \wedge n^{-1}\Lambda(s)] + \sum_{s>m^\circ_n} |\theta^\circ(s)|^2 [1 \wedge n^{-1}\Lambda(s)] \\ &\leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2} n^{-1} m^\circ_n \Lambda_\circ(m^\circ_n) + \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\, \mathfrak{b}^2_{m^\circ_n}(\theta^\circ) \leqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2} \mathcal{R}^{m^\circ_n}_n(\theta^\circ, \Lambda)\end{aligned} \quad (1.14)$$

which in turn implies $\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m^\circ_n}}) \leqslant 4\|\theta^\circ_{\underline{0}}\|^2_{l^2} \mathcal{R}^\circ_n(\theta^\circ, \Lambda)$. In other words, the estimation of the unknown operator T is negligible whenever $n \leqslant n_\lambda$.

Considering then the behaviour of the oracle rate, we have the following results. We note that in case (p) $\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ, \Lambda) \leqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2} \Lambda_+(K) n_\lambda^{-1}$ and hence

$$\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m^\circ_n}}, \theta^\circ, \Lambda) \leqslant \{[1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2}]\{K\Lambda_\circ(K)n^{-1} + \Lambda_+(K)n_\lambda^{-1}\} \quad (1.15)$$

for all $n_\lambda \in \mathbb{N}$ and $n \geqslant n_{\theta^\circ}$ with n_{θ° as in Remark 1.3.1. In other words the rate is parametric in both the ε-sample size n_λ and the Y-sample size n. Thereby, the additional estimation of λ is negligible whenever $n_\lambda \geqslant n$. In the opposite case (np), it is obviously of interest to characterise the minimal size n_λ of the additional sample from ε needed to attain the same rate as in case of a known operator. We carried this discussion in Num. discussion 1.3.2.

On the other hand, if one is interested in the maximal risk, we have the following results. For all $n_\lambda \in \mathbb{N}$ holds $\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)} \mathcal{R}^\circ_n(\theta^\circ, \Lambda) \leqslant r^2 \mathcal{R}^\star_{n_\lambda}(\mathfrak{a}, \Lambda)$, since for all $\theta^\circ \in \Theta(\mathfrak{a}, r)$

$$\mathcal{R}^\star_{n_\lambda}(\theta^\circ, \Lambda) = \sum_{s\in\mathbb{N}^\star} |\theta^\circ(s)|^2 [1 \wedge \Lambda(s)/n_\lambda] \leqslant \max_{s\in\mathbb{N}} \{\mathfrak{a}(s)^2 \min(1, \Lambda(s)/n_\lambda)\} \|\theta^\circ\|^2_{1/\mathfrak{a}}. \quad (1.16)$$

It follows for all $m, n, n_\lambda \in \mathbb{N}$ immediately that

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m}}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant (r^2+8)\mathcal{R}_n^m(\mathfrak{a}, \Lambda) + 16r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda). \tag{1.17}$$

The upper bound in the last display depends on the dimension parameter m and hence by choosing an optimal value $m_n^\star$ as in (1.5) the upper bound will be minimised, that is

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m_n^\star}}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant (r^2+8)\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) + 16r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda). \tag{1.18}$$

Since the operator T is not known, it is natural to consider a maximal risk also over a class for λ characterising the behaviour of $(\Lambda(s) = |\lambda(s)|^{-2})_{s\in\mathbb{N}}$, precisely $\mathcal{E}_\mathfrak{e}^d := \{\lambda \in l_2 : d^{-2} \leqslant \mathfrak{e}_s|\lambda(s)|^2 = \mathfrak{e}_s\Lambda(s)^{-1} \leqslant d^2, \ \forall s \in \mathbb{N}^\star\}$. We shall note that for all $m \in \mathbb{N}$ and any $\lambda \in \mathcal{E}_\mathfrak{e}^d$, $d^{-2} \leqslant \Lambda_+(m)/\mathfrak{e}_{(m)} \leqslant d^2$, $d^{-2} \leqslant \Lambda_\circ(m)/\overline{\mathfrak{e}}_m \leqslant d^2$. Setting for all $n, n_\lambda \in \mathbb{N}$

$$\begin{aligned} \mathcal{R}_n^m(\mathfrak{a}, \mathfrak{e}) &:= [\mathfrak{a}(m)^2 \vee m\overline{\mathfrak{e}}_m n^{-1}], \qquad m_n^\star(\mathfrak{a}, \mathfrak{e}) := \arg\min\{\mathcal{R}_n^m(\mathfrak{a}, \mathfrak{e}), m \in \mathbb{N}\}, \\ \mathcal{R}_n^\star(\mathfrak{a}, \mathfrak{e}) &:= \mathcal{R}_n^{m_n^\star(\mathfrak{a},\mathfrak{e})}(\mathfrak{a}, \mathfrak{e}) = \min\{\mathcal{R}_n^m(\mathfrak{a}, \mathfrak{e}), m \in \mathbb{N}\} \quad \textit{and} \\ \mathcal{R}_n^\star(\mathfrak{a}, \mathfrak{e}) &:= \max\{\mathfrak{a}(s)\min(1, \mathfrak{e}_s/n_\lambda), s \in \mathbb{N}\}. \end{aligned} \tag{1.19}$$

we have

$$\begin{aligned} \mathcal{R}_n^\star(\mathfrak{a}, \Lambda) &= \min_{m\in\mathbb{N}}\{[\mathfrak{a}(m) \vee m\Lambda_\circ(m)n^{-1}]\} \leqslant d^2 \min_{m\in\mathbb{N}}\{[\mathfrak{a}(m) \vee m\overline{\mathfrak{e}}_m n^{-1}]\} \leqslant d^2\mathcal{R}_n^m(\mathfrak{a}, \mathfrak{e}) \\ \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) &= \max_{s\in\mathbb{N}}\{\mathfrak{a}(s)^2[1 \wedge \Lambda(s)/n_\lambda]\} \leqslant d^2\mathcal{R}_n^\star(\mathfrak{a}, \mathfrak{e}). \end{aligned} \tag{1.20}$$

It follows for all $m, n \in \mathbb{N}$ immediately that

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m}}, \Theta(\mathfrak{a}, r), \mathcal{E}_\mathfrak{e}^d\big) \leqslant (r^2+8d^2)\mathcal{R}_n^\star(\mathfrak{a}, \mathfrak{e}) + 8(\mathcal{C}_4+1)d^2r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \mathfrak{e}). \tag{1.21}$$

Johannes and Schwarz (2013a) have shown that $\inf_{\widehat{\theta}} \mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}, \Theta(\mathfrak{a}, r), \mathcal{E}_\mathfrak{e}^d\big)$, where the infimum is taken over all possible estimators $\widehat{\theta}$ of θ°, is up to a constant bounded from below by $\mathcal{R}_n^\star(\mathfrak{a}, \mathfrak{e}) \vee \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \mathfrak{e})$. Consequently, the rate $(\mathcal{R}_n^\star(\mathfrak{a}, \mathfrak{e}) \vee \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \mathfrak{e}))_{n\in\mathbb{N}}$, the dimension parameters $(m_n^\star(\mathfrak{a}))_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m_n^\star(\mathfrak{a})}})_{n\in\mathbb{N}}$, respectively, is a minimax rate, a minimax dimension and minimax optimal (up to a constant). □

1.6 Circular density deconvolution

Let X and ε be circular random variables (that is to say, taking values in the unit circle), and we describe their position by a measure of angle taking values in $[0, 1[$; we denote $\mathbb{P}_X$ and $\mathbb{P}_\varepsilon$ the respective distributions of these measures of angle. Assume that $\mathbb{P}_X$ and $\mathbb{P}_\varepsilon$ admit respective densities f and h with respect to the Lebesgue measure on $[0, 1]$, denoted μ and we denote $\mathcal{B}$ the Borel σ-algebra on $[0, 1]$.

Definition 35 Modular addition

From now on we denote by $\square$ the modular addition on $[0, 1[$. That is to say, for any x and y in $[0, 1[$, $x\square y = x + y - \lfloor x + y\rfloor$. □

We want to estimate f while observing replications of the random variable $Y = X\Box\varepsilon$ which distribution we denote $\mathbb{P}_Y$. One would notice that $\mathbb{P}_Y$ is given, for any A in $\mathcal{B}$, by $\mathbb{P}_Y(A) = (\mathbb{P}_X \star \mathbb{P}_\varepsilon)(A) = \int_{[0,1[} \int_{[0,1[} \mathbb{1}_A(x\Box s)\ \mathrm{d}\mathbb{P}_X(x)\ \mathrm{d}\mathbb{P}_\varepsilon(s)$. Moreover, $\mathbb{P}_Y$ also admits a density with respect to the Lebesgue measure, denoted g and for any y in $[0,1]$, it is given by $g(y) = (f \star h)(y) = \int_0^1 f(y\Box(-s))h(s)\ \mathrm{d}\mu(s)$. Indeed, for any μ-measurable and μ-almost surely bounded function t, we have

$$\begin{aligned}
\mathbb{E}\left[t(Y)\right] =&\mathbb{E}\left[t(X\Box\varepsilon)\right] = \int_0^1 \int_0^1 t(x\Box s)\ \mathrm{d}\mathbb{P}_X(x)\ \mathrm{d}\mathbb{P}_\varepsilon(s) \\
&= \int_0^1 \int_0^1 t(y)\ \mathrm{d}\mathbb{P}_X(y\Box(-s))\ \mathrm{d}\mathbb{P}_\varepsilon(s) = \int_0^1 t(y) \int_0^1\ \mathrm{d}\mathbb{P}_\varepsilon(s)\ \mathrm{d}\mathbb{P}_X(y\Box(-s)) \\
&= \int_0^1 t(y) \int_0^1 f(y\Box(-s))h(s)\ \mathrm{d}\mu(s)\ \mathrm{d}\mu(y).
\end{aligned}$$

Keeping in mind the notations introduced previously, we have Ξ is the space of square integrable complex valued functions defined on $[0,1[$, and T is the operator which associated to any such function t the function given by $t \star h$. We equip Ξ with the usual internal addition; outer product; and scalar product. That is to say, for any two functions x and y from $[0,1]$ to $\mathbb{C}$, we have $x + y : t \mapsto x(t) + y(t)$; with any a in $\mathbb{C}$, $a \cdot x : t \mapsto a \cdot x(t)$; and finally $\langle x|y\rangle_{L^2} = \int_{[0,1]} x(t)\overline{y(t)}\ \mathrm{d}t$, keeping in mind that for any complex number z, we denote by $\overline{z}$ its conjugated complex number. We hence will use the complex trigonometric orthonormal basis and the associated Fourier transform.

Notation 5 Let be the orthonormal complex trigonometric basis of Ξ, for any s in $\mathbb{Z}$ we have:

$$e_s : [0,1] \to \mathbb{C}, \quad t \mapsto \exp[-2\imath\pi s x].$$

Denoting $\mathcal{M}([0,1[)$ the space of measures on $[0,1]$, we define $\mathcal{F}$, the Fourier transform operator on this set:

$$\mathcal{F} : \mathcal{M}([0,1[) \to \mathbb{C}(\mathbb{Z}), \quad \nu \mapsto [\mu] := \mathcal{F}(\mu) = \left(s \mapsto \int_0^1 e_s(t)\ \mathrm{d}\nu(t)\right).$$

In particular, if $\mathbb{P}$ is a probability measure and Z is a random variable with distribution $\mathbb{P}$, we have for any s in $\mathbb{Z}$, that $[\mathbb{P}](s) = \mathbb{E}[e_s(Z)]$.
We use the same notation for the Fourier transform on Ξ:

$$\mathcal{F} : \Xi \to \mathbb{C}^{\mathbb{Z}}, \quad x \mapsto [x] := \mathcal{F}(x) = \left(s \mapsto \int_0^1 e_s(t)x(t)\ \mathrm{d}t\right).$$

In particular, if x is a density associated with a probability distribution $\mathbb{P}$, their Fourier transforms coincide. $\square$

As a consequence, Θ will be the space of $\mathbb{Z}$-indexed, $\mathbb{C}$-valued, square summable sequences. It is equipped with the usual operation, for any $[x]$ in Θ, we have $\overline{[x]} : s \mapsto \overline{[x](s)}$; with $[y]$ in Θ, we have $[x]+[y] : s \mapsto [x](s)+[y](s)$, as well as, $[x]\cdot[y] : s \mapsto [x](s)\cdot[y](s)$; in addition, with a in $\mathbb{C}$, $a \cdot [x] : s \mapsto a \cdot [x](s)$; and finally we have $\langle [x]|[y]\rangle_{l^2} = \sum_{s\in\mathbb{Z}} [x](s)\overline{[y]}(s)$.

REMARK 1.6.1 *It is convenient to note that for any t_1 and t_2 in $[0,1[$ and s in $\mathbb{Z}$, we have $e_s(t_1 \Box t_2) = e_s(t_1)e_s(t_2)$, due to the periodicity of the complex exponential function. Hence, for two functions x and y of Ξ, we hace $\mathcal{F}(x \star y) = \mathcal{F}(x) \cdot \mathcal{F}(y)$.*

Let us recall the following notations.

NOTATION 6 We denote $\theta^\circ := \mathcal{F}(f); \lambda := \mathcal{F}(h); \phi := \mathcal{F}(g) = \theta^\circ \cdot \lambda$.

Notice that, due to the fact that f, g and h are densities associated with some probability distributions, we know that they belong to a specific subspace of Ξ, more precisely, the subspace of positive-valued functions with integral 1. It is interesting to wonder what is the image set of this subset of Ξ with respect to $\mathcal{F}$. Let's introduce the subspace of $\mathbb{C}^{\mathbb{Z}}$ of so-called positive (semi-)definitive sequences.

DEFINITION 36 A $\mathbb{C}$-valued sequence $[x]$ indexed by $\mathbb{Z}$ is positive (semi-)definite iff, for any natural integer q and vector $\{s_1, \dots, s_q\}$ with entries in $\mathbb{Z}$, the Toeplitz matrix $A = (a_{i,j})_{(i,j)\in[\![1,q]\!]^2}$ with $a_{i,j}$ defined by $[x](s_i - s_j)$ is positive (semi-)definite.
In particular, this requires that $[x](s) = \overline{[x]}(-s)$, $[x](0) > 0$, and for all s, $[x](s) \leqslant [x](0)$. □

Then, by denoting $\mathcal{S}^+(\mathbb{Z})$ the set of all positive definite, complex valued, sequences $[x]$ indexed by $\mathbb{Z}$ with $[x](0) = 1$, we formulate Herglotz's representation theorem, which is a special case of Bochner's theorem.

THEOREM 1.6.1.
A function $[x]$ from $\mathbb{Z}$ to $\mathbb{C}$ with $[x](0) = 1$ is semi-definite positive iff there exist μ in $\mathcal{M}([0,1[)$ such that for all s in $\mathbb{Z}$, we have $[x](s) = [\mu](s)$. □

However, notice that a semi-definite positive sequence needs not to be square summable, and hence the associated measure does not always admit a density (and in particular not a square summable one) with respect to the Lebesgue measure. Nonetheless, by Plancherel theorem, any sequence $[x]$ is square summable if and only if its inverse Fourier transform also is.
To sum up, given a probability measure $\mathbb{P}$ on $[0,1]$ admitting a square integrable density p with respect to the Lebesgue measure; their Fourier transforms $[\mathbb{P}]$ and $[p]$ have the following properties:

$$\begin{aligned} [p] &= [\mathbb{P}]; & \sum_{s\in\mathbb{Z}} |[p](s)|^2 < \infty \quad & p \text{ is square summable;} \\ [p](s) &= \overline{[p]}(-s) \quad p \text{ is real valued;} & [p](0) = 1 \quad & p \text{ integrates to 1;} \end{aligned}$$

and $[p]$ positive semi-definitive implies the positivity of p.

We now consider the implications of Assumption 3, Assumption 4 in this model.
Under Assumption 3, we assume that $\mathbb{P}_\varepsilon$, and hence h and λ, are known. The $\mathbb{Z}$-indexed, $[0,1]$-valued stochastic process $Y = (Y_p)_{p\in\mathbb{Z}}$ is strictly stationary with $Y_0 \sim \mathbb{P}_Y$ and hence, for any s in $\mathbb{Z}$, we have $\mathbb{E}[e_s(Y_0)] = \phi(s) = \theta^\circ(s)\lambda(s)$. As we observe $Y^n = (Y_p)_{p\in[\![1,n]\!]}$, we define, for any s in $\mathbb{Z}$, $\phi_n(s) = \sum_{p=1}^n e_s(Y_p)/n$ and $\theta_n(s) = \phi_n(s)/\lambda(s)$.

Under Assumption 4, $\mathbb{P}_\varepsilon$ is not known, and hence neither are h and λ. We hence consider two $\mathbb{Z}$-indexed, $[0,1]$-valued stochastic processes $Y = (Y_p)_{p\in\mathbb{Z}}$ and $\varepsilon = (\varepsilon_p)_{p\in\mathbb{Z}}$ which are strictly stationary with $Y_0 \sim \mathbb{P}_Y$ and $\varepsilon_0 \sim \mathbb{P}_\varepsilon$ and hence, for any s in $\mathbb{Z}$, we have $\mathbb{E}[e_s(Y_0)] = \phi(s) = \theta^\circ(s)\lambda(s)$ and $\mathbb{E}[e_s(\varepsilon_0)] = \lambda(s)$. We observe the sub-vectors $Y^n = (Y_p)_{p\in[\![1,n]\!]}$ and $\varepsilon^{n_\lambda} = (\varepsilon_p)_{p\in[\![1,n_\lambda]\!]}$ and we define $\phi_n(s) = n^{-1}\sum_{p=1}^{n} e_s(Y_p)$, $\lambda_{n_\lambda}(s) = n_\lambda^{-1}\sum_{p=1}^{n_\lambda} e_s(\varepsilon_p)$, $\lambda^+_{n_\lambda}(s) = \mathbb{1}_{\{|\lambda_{n_\lambda}(s)|^2 > (n_\lambda)^{-1}\}}\lambda_{n_\lambda}(s)^{-1}$ and $\theta_{n,n_\lambda}(s) = \phi_n(s)\lambda_{n_\lambda}(s)^+$.
We now separate the study of the convergence rates of projection estimators and of minimax rates depending on the assumptions. In this perspective it is useful to remind the following notations.

Notation 7 For any s in $\mathbb{Z}$ we define, $\Lambda(s) = |\lambda(s)|^{-2}$, we obviously have $\Lambda(s) = \Lambda(-s)$ and $\Lambda(0) = 1$. In addition, for any m in $\mathbb{N}$, we defined $\Lambda_+(m) = \max_{|s|\leqslant m}\{\Lambda(s)\}$, $\Lambda_\circ(m) = m^{-1}\sum_{0<s\leqslant m}\Lambda(s)$, and $\mathfrak{b}^2_m(\theta^\circ) = \sum_{|s|<m}|\theta^\circ|^2$.

Before moving on to the concrete study of the convergence rates for this model, let us illustrate in 1.4 the impact of the noise density on the observation density. We see that a faster decay of the Fourier coefficients (top left of each panel) translate to a smoother density for the noise (top right panel) and how it influences the observation density (bottom right panel), while the density of interest (bottom left panel) remains unchanged. It is obvious that the convolution operator as a neutral element (the Dirac distribution δ_0), which corresponds to the direct problem case, and an absorbing element (the uniform distribution) where the problem cannot be solved.
The practical implications of this phenomenon can be seen when comparing fig. 1.5 and fig. 1.6. In fig. 1.5 we can compare the projection estimator with threshold values 1, 8, 16, and 24 to the true parameter while observing a sample from the direct problem (the noise density is the Dirac distribution). We can see there that while the estimate with threshold parameter 8 is the closest to the truth, the degradation with values 16 and 24 does not seem too bad. In fig. 1.6, the same objects are represented, however, the sample, which as the same size as in fig. 1.5, is from the inverse problem where the noise density is super-smooth. We see that, the estimation with threshold 8 is not as good as in the direct case but also the degradation when the parameter value is larger is way worth.

1.6.1 Known noise density, independent observations process

We place ourselves under Assumption 3 and Assumption 5. We hence observe an iid. n-sample $Y_1,\dots,Y_n$ from $g = f \star h$. Given an estimator $\widehat{\theta}$ of $\theta^\circ \in l_2$ based on the observations we measure its accuracy by a quadratic risk, that is, $\mathbb{E}\|\widehat{\theta} - \theta^\circ\|^2_{l^2}$. Keep in mind that throughout the thesis we assume that $|\lambda(s)| > 0$ holds for all $s \in \mathbb{Z}$. Considering $\Lambda = (\Lambda(s))_{s\in\mathbb{N}}$ with $\Lambda(s) := |\lambda(s)|^{-2}$ for $s \in \mathbb{N}$, we set $\Lambda_+(m) = \max\{\Lambda(s), s \in [\![1,m]\!]\}$ and $\Lambda_\circ(m) = \frac{1}{m}\sum_{s=1}^{m}\Lambda(s)$.
Notice that due to $\mathbb{E}[\|\theta_n - \theta^\circ\|^2_{l^2}] + n^{-1}\|\theta^\circ_{\underline{0}}\|^2_{l_2} = n^{-1}\sum_{-m\leqslant s\leqslant m}\Lambda(s) + \|\theta^\circ_{\underline{0}}\|^2_{l^2}(1+n^{-1})\,\mathfrak{b}^2_m(\theta^\circ)$ together with, for any s in $\mathbb{Z}$, $\Lambda(s) = \Lambda(-s)$, $|\phi(0)| = 1$ and $|\phi(s)| < 1$ for $s \neq 0$ we indeed have

$$\mathbb{E}[\|\theta_n - \theta^\circ\|^2_{l^2}] + n^{-1}\|\theta^\circ_{\underline{0}}\|^2_{l_2} \leqslant 2n^{-1}m\Lambda_\circ(s) + \|\theta^\circ_{\underline{0}}\|^2_{l^2}(1+n^{-1})\,\mathfrak{b}^2_m(\theta^\circ).$$

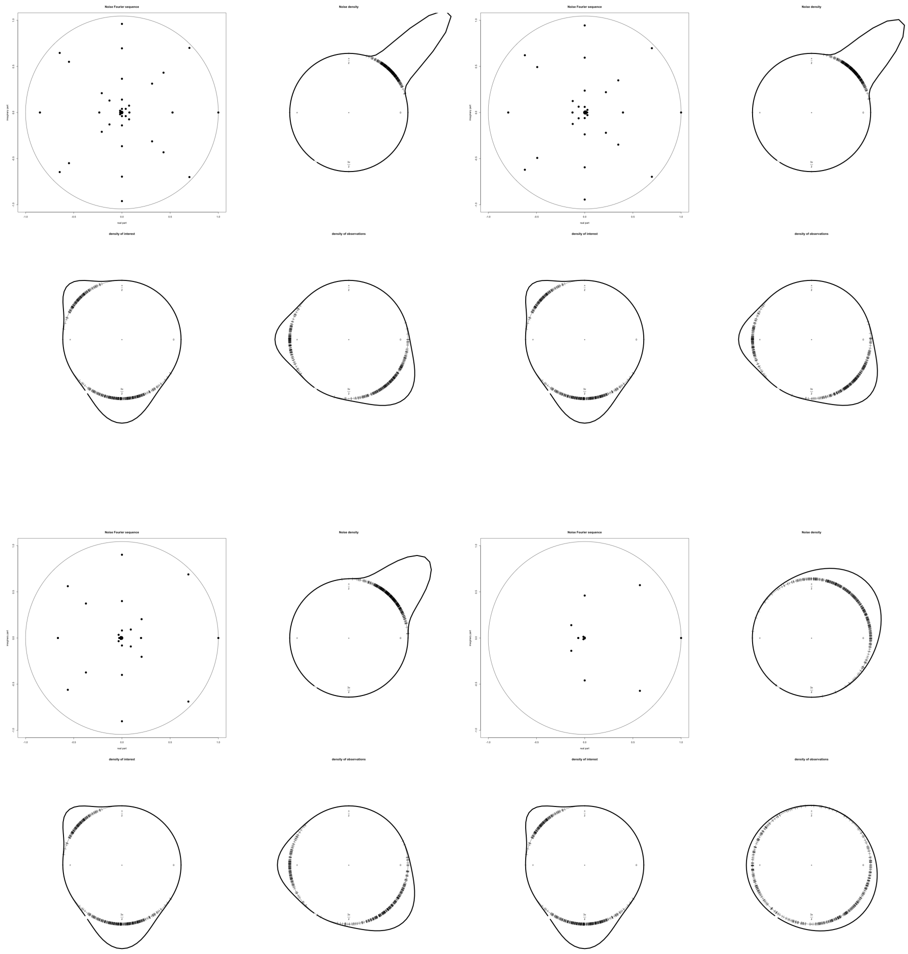

Figure 1.4: Influence of the noise density smoothness on the observation density

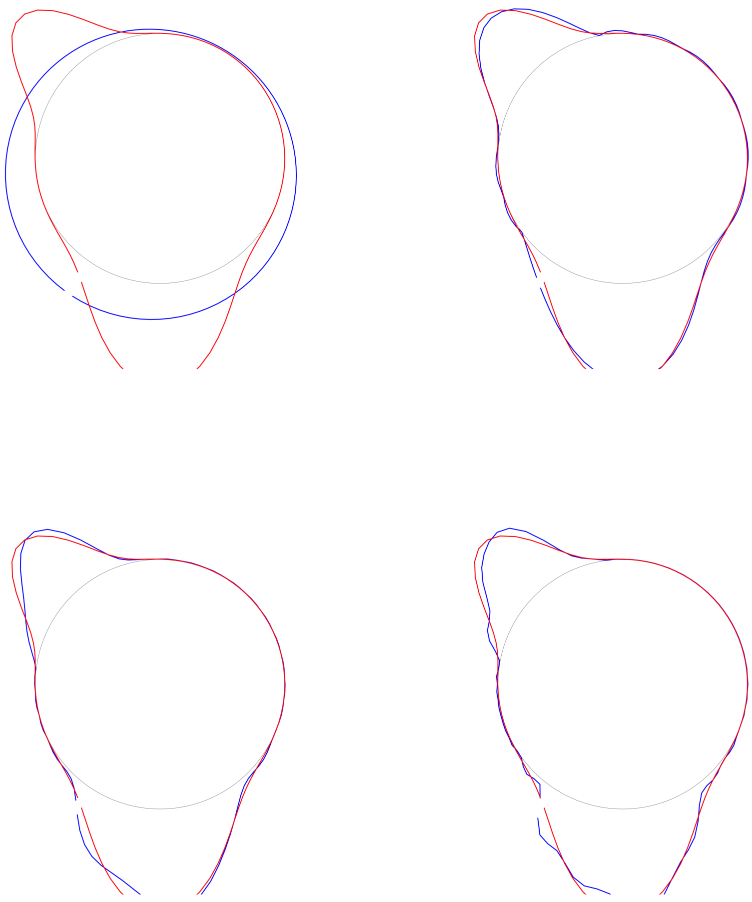

Figure 1.5: Influence of the threshold parameter choice on the estimation in the direct problem case

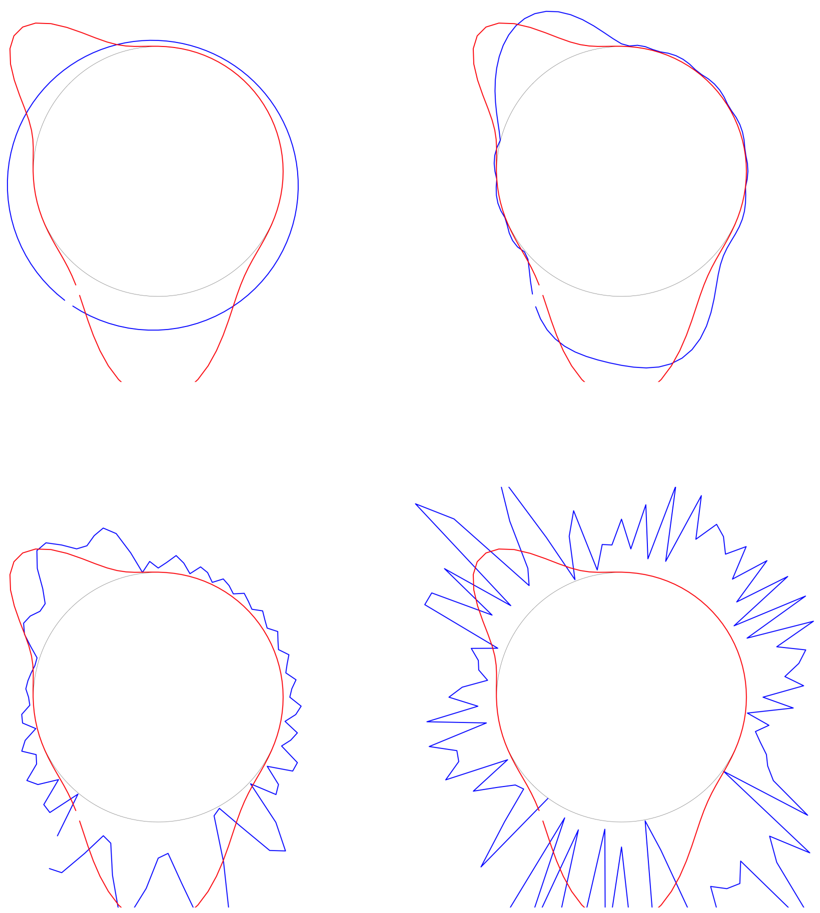

Figure 1.6: Influence of the threshold parameter choice on the estimation in the severely ill-posed problem case

Finally, since $(\lambda(s))_{s\in\mathbb{Z}}$ is a l^2 sequence having only non-zero components bounded by one, i.e., $0 < |\lambda(s)| \leqslant 1$, for all $s \in \mathbb{Z}$, it follows $\lim_{m\to\infty} \Lambda_+(m) = \infty$ and for any diverging sequence $(m_n)_{n\in\mathbb{N}}$ of positive integers, i.e., $\lim_{n\to\infty} m_n = \infty :\Leftrightarrow \forall K > 0 : \exists n_o \in \mathbb{N} : \forall n \geqslant n_o : m_n \geqslant K$, holds $\lim_{m\to\infty} m_n\Lambda_\circ(m_n) = \infty$. Notice that, as in Assumption 9, we have $\mathbb{V}[e_s(Y)] = \mathbb{E}\,|e_s(Y)|^2 - |\,\mathbb{E}[e_s(Y)]|^2 = 1 - |\phi(s)|^2$ which is bounded from above by 1. The analysis we carried out previously hence is still valid and we remind the following definitions.

1.6.1.1 Quadratic risk bounds

The bound we derived in notation 3 depends on the dimension parameter m and hence by selecting an optimal value they will be minimised, which we formulate next. For a sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers with minimal value in a set $A \subset \mathbb{N}$ we set $\arg\min\{a_n, n \in A\} := \min\{m \in A : a_m \leqslant a_n,\ \forall n \in A\}$. For all $n \in \mathbb{N}$ we define

$$\mathcal{R}_n^m(\theta^\circ, \Lambda) := [\mathfrak{b}_m^2(\theta^\circ) \vee m\Lambda_\circ(m)n^{-1}] := \max\left(\mathfrak{b}_m^2(\theta^\circ), m\Lambda_\circ(m)n^{-1}\right),$$
$$m_n^\circ := m_n^\circ(\theta^\circ, \Lambda) := \arg\min\{\mathcal{R}_n^m(\theta^\circ, \Lambda), m \in \mathbb{N}\} \quad \text{and}$$
$$\mathcal{R}_n^\circ(\theta^\circ, \Lambda) := \mathcal{R}_n^{m_n^\circ}(\theta^\circ, \Lambda) = \min\{\mathcal{R}_n^m(\theta^\circ, \Lambda), m \in \mathbb{N}\}.$$

Consequently, the rate $(\mathcal{R}_n^\circ(\theta^\circ, \Lambda))_{n\in\mathbb{N}}$, the dimension parameters $(m_n^\circ)_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m_n^\circ}})_{n\in\mathbb{N}}$, respectively, is an oracle rate, an oracle dimension and oracle optimal (up to a constant).

Remark 1.6.2 *We shall emphasise that $\mathcal{R}_n^\circ(\theta^\circ, \Lambda) \geqslant n^{-1}$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty} \mathcal{R}_n^\circ(\theta^\circ, \Lambda) = 0$. Observe that for all $\delta > 0$ there exists $m_\delta \in \mathbb{N}$ and $n_\delta \in \mathbb{N}$ such that for all $n \geqslant n_\delta$ holds $\mathfrak{b}_{m_\delta}^2(\theta^\circ) \leqslant \delta$ and $m_\delta\Lambda_\circ(m_\delta)n^{-1} \leqslant \delta$, and whence $\mathcal{R}_n^\circ(\theta^\circ, \Lambda) \leqslant \mathcal{R}_n^{m_\delta}(\theta^\circ, \Lambda) \leqslant \delta$. Moreover, we have $m_n^\circ \in [\![1, n]\!]$. Indeed, by construction holds $\mathfrak{b}_n^2(\theta^\circ) \leqslant 1 < (n+1)n^{-1} \leqslant (n+1)\Lambda_\circ(n+1)n^{-1}$, and hence $\mathcal{R}_n^n(\theta^\circ, \Lambda) < \mathcal{R}_n^m(\theta^\circ, \Lambda)$ for all $m \in [\![n+1, \infty[\![$ which in turn implies the claim $m_n^\circ \in [\![1, n]\!]$. Obviously, it follows thus $\mathcal{R}_n^\circ(\theta^\circ, \Lambda) = \min\{\mathcal{R}_n^m(\theta^\circ, \Lambda), m \in [\![1, n]\!]\}$ for all $n \in \mathbb{N}$. We shall use those elementary findings in the sequel without further reference. The sequence $\mathcal{R}_n^\circ(\theta, \lambda)$ is then an exact oracle convergence rate and the projection estimator $\theta_{n,\overline{m_n^\circ}}$ is an oracle optimal estimator.* □

Remark 1.6.3 *In case (p), the oracle rate is parametric, that is $\mathcal{R}_n^\circ(\theta^\circ, \Lambda) \approx n^{-1}$. More precisely, if $\theta^\circ = 0$ then for each $m \in \mathbb{N}$, $\mathbb{E}\|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2 = 2m\Lambda_\circ(m)n^{-1}$, and hence $m_n^\circ = 1$ and $\mathcal{R}_n^\circ(\theta^\circ, \Lambda) = 2\Lambda_\circ(1)n^{-1} \sim n^{-1}$. Otherwise if there is $K \in \mathbb{N}$ with $\mathfrak{b}_{K-1}(\theta^\circ) > 0$ and $\mathfrak{b}_K(\theta^\circ) = 0$, then setting $n_{\theta^\circ} := \frac{K\Lambda_\circ(K)}{\mathfrak{b}_{K-1}^2(\theta^\circ)}$, for all $n \geqslant n_{\theta^\circ}$ holds $\mathfrak{b}_{K-1}^2(\theta^\circ) > K\Lambda_\circ(K)n^{-1}$, and hence $m_n^\circ = K$ and $\mathcal{R}_n^\circ(\theta^\circ, \Lambda) = K\Lambda_\circ(K)n^{-1} \sim n^{-1}$. On the other hand side, in case (np) the oracle rate is non-parametric, more precisely, it holds $\lim_{n\to\infty} n\mathcal{R}_n^\circ(\theta^\circ, \Lambda) = \infty$. Indeed, since $\mathfrak{b}_{m_n^\circ}^2(\theta^\circ) \leqslant \mathcal{R}_n^\circ(\theta^\circ, \Lambda) = \mathcal{R}_n^\circ(\theta^\circ, \Lambda) \in \mathfrak{o}_n(1)$ follows $m_n^\circ \to \infty$ and hence $m_n^\circ\Lambda_\circ(m_n^\circ) \to \infty$ which implies the claim because $n\mathcal{R}_n^\circ(\theta^\circ, \Lambda) \geqslant m_n^\circ\Lambda_\circ(m_n^\circ)$.*

1.6.1.2 Maximal risk bounds

We may emphasise that for all $m \in \mathbb{N}^\star$ and any $\theta^\circ \in \Theta(\mathfrak{a}, r)$, $\|\theta^\circ_{\underline{0}}\|^2_{l^2}\, \mathfrak{b}^2_m(\theta^\circ) = \|\theta^\circ_{\underline{m}}\|^2_{l^2} = \sum_{|s|>m}(\mathfrak{a}(s)^2/\mathfrak{a}(s)^2)\theta^\circ(s)^2 \leqslant \mathfrak{a}(m)^2\|\theta^\circ_{\underline{m}}\|^2_{1/\mathfrak{a}} \leqslant \mathfrak{a}(m)^2 r^2$ which we use in the sequel without further reference. It follows for all $m, n \in \mathbb{N}$ that

$$\begin{aligned}\mathcal{R}_n\big(\theta_{n,\overline{m}}, \Theta(\mathfrak{a}, r), \Lambda\big) := \sup\big\{\mathcal{R}_n\big(\theta_{n,\overline{m}}, \theta^\circ, \Lambda\big), \theta^\circ \in \Theta(\mathfrak{a}, r)\big\}&\\ \leqslant (2 + r^2)\max\big(\mathfrak{a}(m)^2, m\Lambda_\circ(m)n^{-1}\big).&\end{aligned} \tag{1.22}$$

The upper bound in the last display depends on the dimension parameter m and hence by choosing an optimal value $m_n^\star$ the upper bound will be minimised which we formulate next. For all $n \in \mathbb{N}$ we define

$$\begin{aligned}\mathcal{R}_n^m(\mathfrak{a}, \Lambda) &:= [\mathfrak{a}(m)^2 \vee m\Lambda_\circ(m)n^{-1}] := \max\big(\mathfrak{a}(m)^2, m\Lambda_\circ(m)n^{-1}\big),\\ m_n^\star(\mathfrak{a}) &:= m_n^\star(\mathfrak{a}, \Lambda) := \arg\min\{\mathcal{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\} \quad \text{and}\\ \mathcal{R}_n^\star(\mathfrak{a}, \Lambda) &:= \mathcal{R}_n^{m_n^\star(\mathfrak{a})}(\mathfrak{a}, \Lambda) = \min\{\mathcal{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\}.\end{aligned} \tag{1.23}$$

From (1.4) we deduce that $\mathcal{R}_n\big(\theta_{n,\overline{m_n^\star(\mathfrak{a})}}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant (2 + r^2)\mathcal{R}_n^\star(\mathfrak{a}, \Lambda)$ for all $n \in \mathbb{N}$. On the other hand side, for example, Johannes and Schwarz (2013a) have shown that $\inf_{\widetilde{\theta}} \mathcal{R}_n\big(\widetilde{\theta}, \Theta(\mathfrak{a}, r), \Lambda\big)$, where the infimum is taken over all possible estimators $\widetilde{\theta}$ of θ°, is up to a constant bounded from below by $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda)$. Consequently, the rate $(\mathcal{R}_n^\star(\mathfrak{a}, \Lambda))_{n\in\mathbb{N}}$, the dimension parameters $(m_n^\star(\mathfrak{a}))_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m_n^\star(\mathfrak{a})}})_{n\in\mathbb{N}}$, respectively, is a minimax rate, a minimax dimension and minimax optimal (up to a constant).

Remark 1.6.4 *By construction it holds $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) \geqslant n^{-1}$ for all $n \in \mathbb{N}$. The following statements can be shown using the same arguments as in Remark 1.3.1 by exploiting that the sequence $\mathfrak{a}$ is assumed to be non-increasing, strictly positive with limit zero and $\mathfrak{a}(1) = 1$. Thereby, we conclude that $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) = \mathfrak{o}_n(1)$ and $n\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) \to \infty$ as well as $m_n^\star(\mathfrak{a}) \in [\![1, n]\!]$ for all $n \in \mathbb{N}$. It follows also that $m_n^\star(\mathfrak{a}) = \arg\min\{\mathcal{R}_n^m(\mathfrak{a}, \Lambda), m \in [\![1, n]\!]\}$ and $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda) = \min\{\mathcal{R}_n^m(\mathfrak{a}, \Lambda), m \in [\![1, n]\!]\}$ for all $n \in \mathbb{N}$. We shall stress that in this situation the rate $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda)$ is non-parametric.* □

1.6.2 Unknown noise density, independent observations process

We place ourselves under Assumption 4 and Assumption 5. We hence observe independent iid. n-sample $Y_1, \dots, Y_n$ from g and iid. n_λ-sample $\varepsilon_1, \dots, \varepsilon_{n_\lambda}$ from h. Note that we define the projection estimators in the following way $\theta_{n,n_\lambda,\overline{m}} := \mathbb{1}_{\{|s|\leqslant m\}}\lambda^+_{n_\lambda}(s)\phi_n(s)$ with $\lambda^+_{n_\lambda}(s) := \lambda_{n_\lambda}(s)^{-1}\mathbb{1}_{\{|\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda\}}$.

Note that the following result is given in Theorem 2.10 of Petrov (1995).

Lemma 1.6.1.
There is a finite numerical constant $\mathcal{C}_4 > 0$ such that for all $s \in \mathbb{Z}$ hold

$$n_\lambda^2\, \mathbb{E}|\lambda(s) - \lambda_{n_\lambda}(s)|^4 \leqslant \mathcal{C}_4. \tag{1.24}$$

Hence, Assumption 9 is also valid in this model and so is the analysis we carried out later. We hence remind here the following definitions.

1.6.2.1 Quadratic risk bounds

Let us remind that we have

$$\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m_n^\circ}}) \leqslant (V_2\mathcal{C} + \|\theta_{\underline{0}}^\circ\|_{l^2}^2)\mathcal{R}_n^\circ(\theta^\circ,\Lambda) + 2\mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)$$

We note that $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 = 0$ implies $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) = 0$, while for $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 > 0$ holds $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) \geqslant \sum_{s:\Lambda(s)>n_\lambda}|\theta^\circ(s)|^2 + n_\lambda^{-1}\sum_{s:\Lambda(s)\leqslant n_\lambda}|\theta^\circ(s)|^2 \geqslant n_\lambda^{-1}\sum_{s\in\mathbb{N}}|\theta^\circ(s)|^2 = \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 n_\lambda^{-1}$, thereby whenever $\theta^\circ \neq 0$ any additional term of order $n^{-1} + n_\lambda^{-1}$ is negligible with respect to the rate $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) + \mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)$, since $\mathcal{R}_n^\circ(\theta^\circ,\Lambda) \geqslant n^{-1}$, which we will use below without further reference. We shall emphasise that in case $n = n_\lambda$ it holds

$$\begin{aligned}
\mathcal{R}_n^\dagger(\theta^\circ,\Lambda) &= \sum_{s\in\mathbb{F}_{m_n^\circ}}|\theta^\circ(s)|^2[1\wedge n^{-1}\Lambda(s)] + \sum_{s\in\mathbb{F}_{m_n^\circ}^c}|\theta^\circ(s)|^2[1\wedge n^{-1}\Lambda(s)] \\
&\leqslant \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 n^{-1}m_n^\circ\Lambda_\circ(m_n^\circ) + \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_n^\circ}^2(\theta^\circ) \leqslant \|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathcal{R}_n^{m_n^\circ}(\theta^\circ,\Lambda) \qquad (1.25)
\end{aligned}$$

which in turn implies $\mathcal{R}_{n,n_\lambda}(\theta_{n,n_\lambda,\overline{m_n^\circ}}) \leqslant (V_2\mathcal{C} + (1+2\mathcal{C})\|\theta_{\underline{0}}^\circ\|_{l^2}^2)\mathcal{R}_n^\circ(\theta^\circ,\Lambda)$. In other words, the estimation of the unknown operator T is negligible whenever $n \leqslant n_\lambda$.

Remark 1.6.5 *We note that in case (p) $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) \leqslant \|\theta_{\underline{0}}^\circ\|_{l^2}^2\Lambda_+(K)n_\lambda^{-1}$ and hence*

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m_n^\circ}},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}\{[1\vee\|\theta_{\underline{0}}^\circ\|_{l^2}^2]\{K\Lambda_\circ(K)n^{-1} + \Lambda_+(K)n_\lambda^{-1}\} \qquad (1.26)$$

for all $n_\lambda \in \mathbb{N}$ and $n \geqslant n_{\theta^\circ}$ with n_{θ° as in Remark 1.3.1. In other words the rate is parametric in both the ε-sample size n_λ and the Y-sample size n. Thereby, the additional estimation of the error density is negligible whenever $n_\lambda \geqslant n$. In the opposite case (np), it is obviously of interest to characterise the minimal size n_λ of the additional sample from ε needed to attain the same rate as in case of a known error density. Thus, in the next illustration we let the ε-sample size depend on the Y-sample size n as well. □

1.6.2.2 Maximal risk bounds

In the minimax case, for all $n_\lambda \in \mathbb{N}$ we define

$$\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\Lambda) := \max_{s\in\mathbb{N}}\{\mathfrak{a}(s)^2[1\wedge\Lambda(s)/n_\lambda]\}\|\theta^\circ\|_{1/\mathfrak{a}}^2. \qquad (1.27)$$

then for all $n_\lambda \in \mathbb{N}$ holds $\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\mathcal{R}_{n_\lambda}^\star(\theta^\circ,\Lambda) \leqslant r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\Lambda)$, since for all $\theta^\circ \in \Theta(\mathfrak{a},r)$

$$\mathcal{R}_{n_\lambda}^\star(\theta^\circ,\Lambda) = \sum_{s\in\mathbb{N}}|\theta^\circ(s)|^2[1\wedge\Lambda(s)/n_\lambda] \leqslant \max_{s\in\mathbb{N}}\{\mathfrak{a}(s)^2\min(1,\Lambda(s)/n_\lambda)\}\|\theta^\circ\|_{1/\mathfrak{a}}^2. \qquad (1.28)$$

It follows for all $m, n, n_\lambda \in \mathbb{N}$ immediately that

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m}},\Theta(\mathfrak{a},r),\Lambda\big) \leqslant (r^2+8)\mathcal{R}_n^m(\mathfrak{a},\Lambda) + 8(\mathcal{C}_4+1)r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\Lambda). \qquad (1.29)$$

The upper bound in the last display depends on the dimension parameter m and hence by choosing an optimal value $m_n^\star$ as in (1.5) the upper bound will be minimised, that is

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m_n^\star}}, \Theta(\mathfrak{a},r),\Lambda\big) \leqslant (r^2+8)\mathcal{R}_n^\star(\mathfrak{a},\Lambda) + 8(\mathcal{C}_4+1)r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\Lambda). \tag{1.30}$$

Remark 1.6.6 *Since the operator T is not known, it is natural to consider a maximal risk also over a class for λ characterising the behaviour of $(\Lambda(s) = |\lambda(s)|^{-2})_{s\in\mathbb{N}}$, precisely $\mathcal{E}_{\mathfrak{e}}^d := \{\lambda \in l_2 : d^{-2} \leqslant \mathfrak{e}_s|\lambda|s||^2 = \mathfrak{e}_s/\Lambda(|s|) \leqslant d^2,\ \forall s \in \mathbb{N}\} \cap \mathcal{D}$. We shall note that for all $m\in\mathbb{N}$ and any $\lambda \in \mathcal{E}_{\mathfrak{e}}^d$, $d^{-2} \leqslant \Lambda_+(m)/\mathfrak{e}_{(m)} \leqslant d^2$, $d^{-2} \leqslant \Lambda_\circ(m)/\overline{\mathfrak{e}}_m \leqslant d^2$. Setting for all $n, n_\lambda \in \mathbb{N}$*

$$\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e}) := [\mathfrak{a}(m)^2 \vee m\overline{\mathfrak{e}}_m n^{-1}], \qquad m_n^\star(\mathfrak{a},\mathfrak{e}) := \arg\min\{\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e}), m\in\mathbb{N}\},$$
$$\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) := \mathcal{R}_n^{m_n^\star(\mathfrak{a},\mathfrak{e})}(\mathfrak{a},\mathfrak{e}) = \min\{\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e}), m\in\mathbb{N}\} \quad \textit{and}$$
$$\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) := \max\{\mathfrak{a}(s)\min(1,\mathfrak{e}_s/n_\lambda), s\in\mathbb{N}\}. \tag{1.31}$$

we have

$$\mathcal{R}_n^\star(\mathfrak{a},\Lambda) = \min_{m\in\mathbb{N}}\{[\mathfrak{a}(m)\vee m\Lambda_\circ(m)n^{-1}]\} \leqslant d^2\min_{m\in\mathbb{N}}\{[\mathfrak{a}(m)\vee m\overline{\mathfrak{e}}_m n^{-1}]\} \leqslant d^2\mathcal{R}_n^m(\mathfrak{a},\mathfrak{e})$$
$$\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\Lambda) = \max_{s\in\mathbb{N}}\{\mathfrak{a}(s)^2[1\wedge\Lambda(s)/n_\lambda]\} \leqslant d^2\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}). \tag{1.32}$$

It follows for all $m,n\in\mathbb{N}$ immediately that

$$\mathcal{R}_{n,n_\lambda}\big(\theta_{n,n_\lambda,\overline{m}}, \Theta(\mathfrak{a},r),\mathcal{E}_{\mathfrak{e}}^d\big) \leqslant (r^2+8d^2)\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e}) + 8(\mathcal{C}_4+1)d^2r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\mathfrak{e}). \tag{1.33}$$

Johannes and Schwarz (2013a) have shown that $\inf_{\widehat{\theta}}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta},\Theta(\mathfrak{a},r),\mathcal{E}_{\mathfrak{e}}^d\big)$, where the infimum is taken over all possible estimators $\widehat{\theta}$ of θ°, is up to a constant bounded from below by $\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e})\vee\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\mathfrak{e})$. Consequently, the rate $(\mathcal{R}_n^\star(\mathfrak{a},\mathfrak{e})\vee\mathcal{R}_{n_\lambda}^\star(\mathfrak{a},\mathfrak{e}))_{n\in\mathbb{N}}$, the dimension parameters $(m_n^\star(\mathfrak{a}))_{n\in\mathbb{N}}$ and the projection estimators $(\theta_{n,\overline{m_n^\star(\mathfrak{a})}})_{n\in\mathbb{N}}$, respectively, is a minimax rate, a minimax dimension and minimax optimal (up to a constant). □

Chapter 2

Bayesian interpretation of model selection

In this chapter, we consider the family of Bayesian methods described as "Gaussian sieve priors" in section 1.4.2 as well as an adaptive variant of these priors, the hierarchical sieve priors where the threshold parameter is a random variable with a specified prior distribution. We study their behaviour under two asymptotic, respectively described in section 1.4.3 and section 1.4.1.
In section 2.1 we consider the self informative Bayes carrier of Gaussian sieve priors under continuity assumptions for the likelihood and show that its support is contained in the maximum likelihood set. Then, in section 2.2 we show that the distribution of the hyper-parameter in the hierarchical prior contracts around the set of maximisers of a penalised contrast criterion. This section highlights a new link between Bayesian adaptive estimation and the frequentist penalised contrast model selection.
In section 2.3, while considering the noise asymptotic, we line out two strategies of proof which allow to obtain contraction rates. The first relies on posterior moment bounding, it is easy to apply and we give examples where the obtained bounds are optimal, however, it requires analytical expressions for the posterior moments, which are not often available for the sophisticated priors used in non-parametric Bayes methods; the second is specific to the hierarchical sieve prior and is similar to the one used in Johannes et al. (2014), we generalise it to the self informative Bayes carrier where the posterior distribution is supported by a null-measure set. In section 2.4 we apply this strategies to the specific inverse Gaussian sequence space model. Doing so, we obtain exact contraction rate for the (iterated) Gaussian sieve prior using the first scheme of proof; and the iterated hierarchical prior using the second. This yields optimality for sieve priors with properly chosen threshold parameter; as well as for penalised contrast model selection; and for any iterated version of the hierarchical prior we consider.
Finally, we conclude this chapter in section 2.5 with a note about the shape of the posterior mean of the hierarchical prior, motivating the shape of the frequentist estimators we use in chapter 3.

2.1 Iterated Gaussian sieve prior

We consider in this part a statistical model with a functional parameter space as described in section 1.2. We adopt a sieve prior as described in section 1.4.2 and first give interest

to the iteration asymptotic presented in section 1.4.1.
We assume the existence of a function $l : (\Theta, \mathcal{B}) \times (\Theta^n, \mathcal{B}^{\otimes n}) arrow \mathbb{R}$ such that the likelihood with respect to some reference measures $\mathbb{P}^\circ$ is given by: $L(\boldsymbol{\theta}, y^n) \propto \exp[-l(\boldsymbol{\theta}, y^n)]$.
Then, the family of Gaussian sieve priors is indexed by a threshold parameter m in $\mathbb{M} (= \mathbb{N}$ for our examples), and we denote by $\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}}$ the element of this family with index m; moreover, we denote $\boldsymbol{\theta}_{\overline{m}}$ a random variable following this distribution. There exists a reference measure $\mathbb{Q}^\circ$ such that the Gaussian sieve prior with threshold parameter m admits a density of the shape $d\,\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}} / d\,\mathbb{Q}^\circ(\boldsymbol{\theta}) \propto \exp[-(1/2)\sum_{|s|\leqslant m} |\boldsymbol{\theta}(s)|^2] \cdot \prod_{|s|>m} \delta_0(\boldsymbol{\theta}(s))$.
Denote by $\Theta_{\overline{m}}$ the set $\{\theta \in \Theta : \forall s \in \mathbb{F}_{\underline{m}}, \theta(s) = 0\}$. Bayes' theorem gives the following shape for the iterated posterior distribution:

$$(\mathrm{d}\mathbb{P}^{\eta}_{\boldsymbol{\theta}_{\overline{m}}|Y^n} \,/\, \mathrm{d}\mathbb{Q}^\circ)(\boldsymbol{\theta}, Y^n) \propto \exp[-((1/2)\sum_{|s|\leqslant m} |\boldsymbol{\theta}(s)|^2 + \eta l(\boldsymbol{\theta}, y^n))] \cdot \prod_{|s|>m} \delta_0(\boldsymbol{\theta}(s))$$
$$= \frac{\prod_{|s|>m} \delta_0(\boldsymbol{\theta}(s))}{\int_{\Theta_m} \exp[-(1/2)\sum_{|s|\leqslant m}(|\mu(s)|^2 - |\boldsymbol{\theta}(s)|^2)] \exp[-\eta(l(\mu, y^n) - l(\boldsymbol{\theta}, y^n))] \, \mathrm{d}\mathbb{Q}^\circ(\mu)}.$$

We then place ourselves under the assumption of a continuous likelihood to obtain the self informative Bayes carrier.

Assumption 10 Assume that for any m in $\mathbb{M}$ and y in Ξ^n, $\Theta_{\overline{m}} arrow \mathbb{R}_+, \theta \mapsto l(\theta, y^n)$ is continuous. □

The use of a threshold parameter brings us back to the study of a parametric model and we obtain the following result.

Theorem 2.1.1.
Assuming that $\mathbb{M} = \mathbb{N}$ and Assumption 10, the support of the Bayesian carrier is contained in the set of minimisers of $\theta \mapsto l(\theta, y^n)$ under the constraint $\theta \in \Theta_{\overline{m}}$. □

Proof of Theorem 2.1.1.
Let's remind that the definition of continuity gives us:

$$\forall \theta \in \Theta_{\overline{m}}, \forall \varepsilon \in \mathbb{R}^\star_+, \exists \delta \in \mathbb{R}^\star_+ : \forall \mu \in \Theta_{\overline{m}}, \|\mu - \theta\| < \delta arrow |l(\mu, y^n) - l(\theta, y^n)| < \varepsilon.$$

Then, for any B in $\mathcal{B}$ such that $\inf_{\theta \in B} l(\theta, y) > \inf_{\mu \in \Theta_m} l(\mu, y)$, there exist δ in $\mathbb{R}^\star_+$ and a ball $\mathcal{E}$ of Θ_m of radius δ such that, $\sup_{\mu \in \mathcal{E}} l(\mu, y^n) < \inf_{\theta \in B} l(\theta, y^n)$ and hence $\sup_{\mu \in \mathcal{E}} l(\mu, y^n) - \inf_{\theta \in B} l(\theta, y^n) < 0$.
Hence, we can write

$$\begin{aligned}
\mathbb{P}^{\eta}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}(B) = \int_B ((\prod_{|s|>m} \delta_0(\boldsymbol{\theta}(s))) \cdot (\int_{\Theta_{\overline{m}}} &\exp[-(1/2)\sum_{|s|\leqslant m}(|\mu(s)|^2 - |\boldsymbol{\theta}(s)|^2)] \\
&\cdot \exp[-\eta(l(\mu, y^n) - l(\boldsymbol{\theta}, y^n))] \; \mathrm{d}\mathbb{Q}^\circ(\mu))^{-1}) \; \mathrm{d}\mathbb{Q}^\circ(\boldsymbol{\theta}) \\
&\leqslant (\exp[-\eta(\sup_{\mu \in \mathcal{E}} l(\mu, y^n) - \inf_{\mu \in B} l(\mu, y^n))])^{-1} \\
&\cdot \int_B ((\prod_{|s|>m} \delta_0(\boldsymbol{\theta}(s)) \exp[-(1/2)\sum_{|s|\leqslant m} |\boldsymbol{\theta}(s)|^2]) \\
&\qquad \cdot (\int_{\mathcal{E}} \exp[-(1/2)\sum_{|s|\leqslant m} |\mu(s)|^2] \; \mathrm{d}\mathbb{Q}^\circ(\mu))^{-1}) \, \mathrm{d}\boldsymbol{\theta} \to 0.
\end{aligned}$$

□

We have hence showed that under the iteration asymptotic, the posterior distribution contracts itself on maximisers of the likelihood, constrained by $\theta(s) = 0$ for any $|s| > m$. We will see that, while considering the noise asymptotic, the choice of the threshold is determinant for the quality of the estimation. The choice of the threshold for the projection estimators and for sieve priors should be led in a similar fashion, that is, balancing the bias (information lost for values of s greater than m) and the variance (noise incorporated in the estimation for values of s which are smaller than m). As stated previously, the ideal choice of this parameter is however dependent on the parameter of interest and hence not available. It is hence important to inquire adaptive methods for the selection of this parameter. Some methods for the frequentist estimation were outlined in the introduction such as the penalised contrast model selection. In the next section, we introduce the hierarchical sieve prior which consists in modelling the threshold parameter as a random variable. We will show that by selecting the prior distribution for this hyper-parameter properly, the iteration asymptotic gives a Bayesian interpretation to the penalised contrast model selection.

2.2 Adaptivity using a hierarchical prior

We denote $\mathbb{P}_{\boldsymbol{\theta}_{\overline{M}}}$ a so called hierarchical prior distribution, described hereafter, and $\boldsymbol{\theta}_{\overline{M}}$ a random variable following this prior. Define G, a finite element of $\mathbb{M}$ (depending on n), acting as an upper bound for M and $\mathrm{pen} : [\![1, G]\!] \to \mathbb{R}_+$ a so-called penalty function. The threshold parameter noted m for the sieve prior described in the previous section is now a $[\![1, G]\!]$-valued random variable denoted M. We note $\mathbb{P}_M$ the distribution of this parameter. The density of $\mathbb{P}_M$ with respect to the counting measure is given, for any m in $[\![1, G]\!]$, by $\mathbb{P}_M(m) \propto \exp[\mathrm{pen}(m)]\mathbb{1}_{\{m \leqslant G\}}$.

The dependance structure between the different quantities of the model is then the following:

$$\mathbb{P}_{\boldsymbol{\theta}_{\overline{M}}|M=m} = \mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}}; \quad \mathbb{P}_{Y^n|\boldsymbol{\theta},M} = \mathbb{P}_{Y^n|\boldsymbol{\theta}}.$$

The following proposition, giving an expression for the iterated posterior distribution of the threshold parameter, is obtained by direct calculus.

Proposition 2.2.1.
For any m in $\mathbb{M}$, we use the convention $\mathbb{P}^0_{\boldsymbol{\theta}_{\overline{m}}|Y^n} = \mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}}$, define for any η in $\mathbb{N}^\star$, Y^n in Ξ^n, and m in $[\![1, G]\!]$, the following quantity

$$\begin{aligned}\exp[\Upsilon^\eta(Y^n, m)] &:= \int_\Theta L(\boldsymbol{\theta}, Y^n)\,(\,\mathrm{d}\mathbb{P}^{\eta-1}_{\boldsymbol{\theta}_{\overline{m}}|Y^n})/(\,\mathrm{d}\mathbb{Q}^\circ)(m, \boldsymbol{\theta})\,\mathrm{d}\mathbb{Q}^\circ(\boldsymbol{\theta}) \\ &= \int_\Theta \exp\left[-\left((1/2)\sum\nolimits_{|s|\leqslant m} |\boldsymbol{\theta}(s)|^2 + \eta l(\boldsymbol{\theta}, y^n)\right)\right]\,\mathrm{d}\mathbb{Q}^\circ(\boldsymbol{\theta}).\end{aligned}$$

The iterated posterior distribution of the threshold parameter is given, for any m in $[\![1, G]\!]$

and y^n *in* Ξ^n *by:*

$$\begin{aligned}\mathbb{P}^{\eta}_{M|Y^n}(m, y) &\propto \exp\left[\mathrm{pen}(m) + \eta\Upsilon^{\eta}(y^n, m)\right] \mathbb{1}_{\{m\leqslant G\}}\\ &= \left(\sum\nolimits_{k\in[\![1,G]\!]} \exp\left[\eta\left(\Upsilon^{\eta}(y^n, k) - \eta\Upsilon^{\eta}(y^n, m)\right) + (\mathrm{pen}(k) - \mathrm{pen}(m))\right]\right)^{-1} \mathbb{1}_{\{m\leqslant G\}}.\end{aligned}$$

Proof of Proposition 2.2.1

$$\begin{aligned}\mathbb{P}_{M|Y^n}(m, y^n) &\propto (\,\mathrm{d}\mathbb{P}_{M,Y^n} \,/\, \mathrm{d}\mathbb{P}^{\circ})(m, y^n)\\
&\propto \int_{\Theta} (\,\mathrm{d}\mathbb{P}_{M,Y^n,\boldsymbol{\theta}_{\overline{m}}} \,/\, \mathrm{d}\mathbb{P}^{\circ}\; d\,\mathbb{Q}^{\circ})(m, y^n, \theta)\; \mathrm{d}\mathbb{Q}^{\circ}(\theta)\\
&\propto \int_{\Theta} (\,\mathrm{d}\mathbb{P}_{Y^n|M,\boldsymbol{\theta}_{\overline{M}}} \,/\, \mathrm{d}\mathbb{P}^{\circ})(m, y^n, \boldsymbol{\theta}) \cdot (\,\mathrm{d}\mathbb{P}_{M,\boldsymbol{\theta}_{\overline{M}}} \,/\, \mathrm{d}\mathbb{Q}^{\circ})(m, \theta)\; \mathrm{d}\mathbb{Q}^{\circ}(\theta)\\
&\propto \int_{\Theta} (\,\mathrm{d}\mathbb{P}_{Y^n|\boldsymbol{\theta}_{\overline{M}}} \,/\, \mathrm{d}\mathbb{P}^{\circ})(y^n, \theta) \cdot (\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{M}}|M} \,/\, \mathrm{d}\mathbb{Q}^{\circ})(m, \theta) \cdot \mathbb{P}_M(m)\; \mathrm{d}\mathbb{Q}^{\circ}(\boldsymbol{\theta})\\
&\propto \mathbb{P}_M(m) \cdot \int_{\Theta} (\,\mathrm{d}\mathbb{P}_{Y^n|\boldsymbol{\theta}_{\overline{M}}} \,/\, \mathrm{d}\mathbb{P}^{\circ})(y^n, \theta) \cdot (\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}} \,/\, \mathrm{d}\mathbb{Q}^{\circ})(m, \theta) \cdot \mathrm{d}\mathbb{Q}^{\circ}(\boldsymbol{\theta})\\
&= \frac{\mathrm{d}\mathbb{P}_M(m) \cdot \int_{\Theta} (\,\mathrm{d}\mathbb{P}_{Y^n|\boldsymbol{\theta}_{\overline{m}}} \,/\, \mathrm{d}\mathbb{P}^{\circ})(y^n, \boldsymbol{\theta}) \cdot (\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}} \,/\, \mathrm{d}\mathbb{Q}^{\circ})(m, \boldsymbol{\theta}) \cdot \mathrm{d}\mathbb{Q}^{\circ}(\boldsymbol{\theta})}{\sum_{|s|\leqslant G} \mathbb{P}_M(s) \cdot \int_{\Theta} (\,\mathrm{d}\mathbb{P}_{Y^n|\boldsymbol{\theta}_{\overline{m}}} \,/\, \mathrm{d}\mathbb{P}^{\circ})(y^n, \theta) \cdot (\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}} \,/\, \mathrm{d}\mathbb{Q}^{\circ})(s, \boldsymbol{\theta})\; \mathrm{d}\mathbb{Q}^{\circ}(\boldsymbol{\theta})}\\
&= \frac{\exp[\mathrm{pen}(m)] \int_{\Theta_{\overline{m}}} \exp[-(1/2)(2l(y^n, \boldsymbol{\theta}) + \sum_{|s|\leqslant m} |\boldsymbol{\theta}(s)|^2)]\; \mathrm{d}\mathbb{Q}^{\circ}(\boldsymbol{\theta})}{\sum_{|s|\leqslant G} \exp[\mathrm{pen}(s)] \int_{\Theta_{\overline{s}}} \exp[-(1/2)(2l(y^n, \boldsymbol{\theta}) + \sum_{|s'|\leqslant s} |\boldsymbol{\theta}(s')|^2)]\; \mathrm{d}\mathbb{Q}^{\circ}(\boldsymbol{\theta})}.\end{aligned}$$

□

From this expression for the iterated posterior distribution we can deduce the self informative Bayes carrier.

Lemma 2.2.1.
Note $\Upsilon(m) := \lim_{\eta\to\infty} \Upsilon^{\eta}(y^n, m)$.
The support of the self informative Bayes carrier for M *is* $\arg\max_{m\leqslant G}\{\Upsilon(m)\}$. □

Unfortunately, in many practical cases, the choice led by $\arg\max_{m\subset G}\{\Upsilon(y^n, m)\}$ is G itself and leads to inconsistent or suboptimal inference (as we will show later). However, if one allows the prior distribution to depend on η and to take the shape $\exp[-\eta\,\mathrm{pen}(m)]\mathbb{1}_{m\subset G}$, we obtain the following result.

Theorem 2.2.1.
Using the modified prior which depends on η*, the support of the self informative Bayes carrier for the hyper-parameter* M *is* $\arg\max_{m\subset G}\{\Upsilon(m) + \mathrm{pen}(m)\}$. □

Proof of Theorem 2.2.1
For any $m \leqslant G$ such that $\Upsilon(m) - \mathrm{pen}(m) < \max_{k\leqslant G} \Upsilon(k) - \mathrm{pen}(k)$, there exist a value of $\eta_{\circ}$ such that, for any η greater than $\eta_{\circ}$, $\Upsilon^{\eta}(m) + \mathrm{pen}(m) < \max_{k\leqslant G} \Upsilon^{\eta}(k) + \mathrm{pen}(k)$ we

can hence write

$$\begin{aligned}\mathbb{P}^{\eta}_{M|Y^n}(m) &= (\textstyle\sum_{k\leqslant G} \exp\left[\eta\left(\Upsilon^{\eta}(k) - \Upsilon^{\eta}(m) + (\mathrm{pen}(k) - \mathrm{pen}(m))\right)\right])^{-1}\mathbb{1}_{\{m\leqslant G\}}\\ &\leqslant (\exp\left[\eta\left(\max_{k\leqslant G}\left(\Upsilon^{\eta}(k) + \mathrm{pen}(k)\right) - \left(\Upsilon^{\eta}(m) + \mathrm{pen}(m)\right)\right)\right])^{-1}\mathbb{1}_{\{m\leqslant G\}}\\ &\to 0.\end{aligned}$$

As $\sum_{m\in\mathbb{N}} \mathbb{P}^{\eta}_{M|Y^n}(m) = 1$ for any η, we have, thanks to the dominated convergence theorem, that for any subset $\mathbb{G}$ of $[\![1, G]\!]$ which does not intersect with $\arg\max_{k\in[\![1,G]\!]}\{\Upsilon^{\eta}(k) + \mathrm{pen}(k)\}$, $\mathbb{P}^{\eta}_{M|Y^n}(\mathbb{G}) = 0$. □

Now that we determined the posterior distribution for the hyper-parameter, we can compute the posterior distribution for $\boldsymbol{\theta}_{\overline{M}}$ itself.

Proposition 2.2.2.
The iterated posterior distribution is given by:

$$\begin{aligned}(\,\mathrm{d}\mathbb{Q}^{\eta}_{\boldsymbol{\theta}_{\overline{M}}|Y}/\,\mathrm{d}\mathbb{P}^{\circ})(\boldsymbol{\theta}, y) &= \textstyle\sum_{m\leqslant G}(\,\mathrm{d}\mathbb{P}^{\eta}_{\boldsymbol{\theta}_{\overline{M}}|Y}\,/\,\mathrm{d}\mathbb{Q}^{\circ})(\boldsymbol{\theta}, y, m)\cdot(\,\mathrm{d}\mathbb{P}^{\eta}_{M|Y}\,/\,\mathrm{d}\mathbb{P}^{\circ})(m, y)\\ &= \sum_{m\leqslant G}\frac{\exp\left[-\left((1/2)\sum_{|s|\leqslant m}|\boldsymbol{\theta}(s)|^2 + \eta l(\boldsymbol{\theta}, y)\right)\right]\cdot\prod_{\{|s|>m\}}\delta_0(\boldsymbol{\theta}(s))}{\int_{\Theta_{\overline{m}}}\exp\left[-\left((1/2)\sum_{|s|\leqslant m}|\mu(s)|^2 + \eta l(\mu, y)\right)\right]\,\mathrm{d}\mathbb{Q}^{\circ}(\mu)}\\ &\quad\cdot\frac{\exp[\mathrm{pen}(m) + \eta\Upsilon^{\eta}(Y, m)]}{\sum_{j\leqslant G}\exp[\mathrm{pen}(j) + \eta\Upsilon^{\eta}(Y, j)]}\mathbb{1}_{m\subset G}\end{aligned}$$

□

Proof for Proposition 2.2.2

$$\begin{aligned}(\,\mathrm{d}\mathbb{Q}_{\boldsymbol{\theta}_{\overline{M}}|Y}/\,\mathrm{d}\mathbb{P}^{\circ})(\boldsymbol{\theta}, y) &\propto (\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{M}},Y}\,/\,\mathrm{d}\mathbb{Q}^{\circ}\,\mathrm{d}\mathbb{P}^{\circ})(\boldsymbol{\theta}, y)\\ &\propto \textstyle\sum_{m\leqslant G}(\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{M}},Y,M}\,/\,\mathrm{d}\mathbb{Q}^{\circ}\,\mathrm{d}\mathbb{P}^{\circ})(\boldsymbol{\theta}, y, m)\\ &\propto \textstyle\sum_{m\leqslant G}(\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{M}}|Y,M}\,/\,\mathrm{d}\mathbb{Q}^{\circ})(\boldsymbol{\theta}, y, m)\cdot(\,\mathrm{d}\mathbb{P}_{Y,M}\,/\,\mathrm{d}\mathbb{P}^{\circ})(y, m)\\ &\propto \textstyle\sum_{m\leqslant G}(\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}|Y}\,/\,\mathrm{d}\mathbb{Q}^{\circ})(\theta, y, m)\cdot(\,\mathrm{d}\mathbb{P}_{M|Y}\,/\,\mathrm{d}\mathbb{P}^{\circ})(m, y)\cdot(\,\mathrm{d}\mathbb{P}_{Y}\,/\,\mathrm{d}\mathbb{P}^{\circ})(y)\\ &= \textstyle\sum_{m\leqslant G}(\,\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_{\overline{m}}|Y}\,/\,\mathrm{d}\mathbb{Q}^{\circ})(\theta, y, m)(\,\mathrm{d}\mathbb{P}_{M|Y}\,/\,\mathrm{d}\mathbb{P}^{\circ})(m, y).\end{aligned}$$

□

And as a consequence, we can deduce the self informative Bayes carrier.

Theorem 2.2.2.
Denote $\widehat{m} := \arg\max_{m\leqslant G}\{\Upsilon(Y, m) + \mathrm{pen}(m)\}$ *then the support of the self informative Bayes carrier is contained in* $\arg\max_{\theta\in\Theta_m, m\in\widehat{m}}\{-l(\theta, Y)\}$.

We have hence seen in these two first sections investigated the behaviour of the sieve prior and its hierarchical version under the iterative asymptotic and shown that under some mild assumptions, their self informative Bayes carriers correspond to some constrained maximum likelihood estimator and penalised contrast model selection version of it respectively. We should now investigate the behaviour of these (iterated) posteriorii under the noise asymptotic and define hypotheses under which they behave properly.

2.3 Proof strategies for contraction rates

In this section, we depict two proof strategies for contraction rates. They will be used in the next sections to compute contraction rates for sieve and hierarchical sieve priors respectively.
The first proof relies on moment bounding of the random variable $\|\boldsymbol{\theta} - \theta^\circ\|$. The second proof relies on the use of exponential concentration inequalities.

2.3.1 Employing control of posterior moments

In this section we outline a method to prove contraction rates which requires to bound properly some moments of the posterior distribution. We later use this method in the case of the inverse Gaussian sequence space with a sieve prior. Provided that bounds are available for the required moments, this method barely needs any other assumption on the model. Moreover, it appears that, in the example we display here, it leads to the same rate as the frequentist optimal convergence rate without a logarithmic loss as it is often the case with popular methods.
This proof scheme is simple, easily interpretable as a link between the convergence of the posterior mean to the true parameter; the contraction of the posterior distribution around the posterior mean and the contraction of the posterior distribution around the true parameter and it gives optimal contraction rates. However, it is not surprising that this method lacks flexibility and could not be applied with too complex priors, as the hierarchical prior we consider here.
However, we believe that the method could be generalised to wider cases, for example using convergence of distribution in Wasserstein distance implying convergence of moments.

For all this section, Φ_n is the sequence which we want to prove to be a contraction rate; it is in general a function of θ° but we do not make this dependence appear in this section as it has no influence on the proof.
This proof relies on the following simple lemma which will be applied consecutively to control the quantities of interest.

Lemma 2.3.1.
Consider a sequence of $\mathbb{R}_+$-valued random variables $(X_n)_{n\in\mathbb{N}}$ such that, for any n in $\mathbb{N}$, we have $\mathbb{E}[|X_n|^2] < \infty$. If $\max\{\mathbb{E}[X_n], \mathbb{V}[X_n]^{1/2}\} \in \mathcal{O}_n(\Phi_n)$, then for any increasing and unbounded sequence $(c_n)_{n\in\mathbb{N}}$, we have $\lim_{n\to\infty} \mathbb{P}(X_n \geqslant c_n\Phi_n) = 0$. □

Proof of Lemma 2.3.1
Define the sequence of random variables $\mathcal{S}_n := (X_n - \mathbb{E}[X_n])/\mathbb{V}[X_n]^{1/2}$. This is a sequence of random variables with common expectation 0 and variance 1 and, as such,

their distributions form a sequence of tight measures. Hence, for any increasing unbounded sequence c_n and $K_n := \mathbb{E}[X_n] + c_n \mathbb{V}[X_n]^{1/2}$ we can write $\mathbb{P}(X_n \geqslant K_n) = \mathbb{P}\left(S_n \geqslant (K_n - \mathbb{E}[X_n])/(\mathbb{V}[X_n]^{1/2})\right) = \mathbb{P}(S_n \geqslant c_n)$ which tends to 0 as $(S_n)_{n\in\mathbb{N}}$ is tight. □

We will hence use this lemma for the two random variables $\mathbb{E}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]$, and $\mathbb{V}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]^{1/2}$. As a consequence, we consider the following assumption, which has to be checked using the specificities of the model on which one plans to use this method.

Assumption 11 Assume $\max\left\{\mathbb{E}\left[\mathbb{E}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]\right], \mathbb{V}\left[\mathbb{E}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]\right]^{1/2}\right\} \in \mathcal{O}(\Phi_n)$ and $\max\left\{\mathbb{E}\left[\mathbb{V}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]^{1/2}\right], \mathbb{V}\left[\mathbb{V}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]^{1/2}\right]^{1/2}\right\} \in \mathcal{O}(\Phi_n)$. □

Notice that, under Assumption 11, using Lemma 2.3.1 gives for any increasing and unbounded $(c_n)_{n\in\mathbb{N}}$ that $\lim_{n\to\infty} \mathbb{P}\left(\mathbb{E}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|] \geqslant c_n\Phi_n\right) = 0$ and $\lim_{n\to\infty} \mathbb{P}\left(\mathbb{V}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]^{1/2} \geqslant c_n\Phi_n\right) = 0$. This gives us the following theorem.

Theorem 2.3.1.
Under Assumption 11, we have for any increasing unbounded sequence c_n

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta} - \theta^\circ\| > c_n\Phi_n)\right] = 0.$$

Proof of Theorem 2.3.1
Denote $E := \mathbb{E}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]$ and $V := \mathbb{V}_{\boldsymbol{\theta}|Y^n}[\|\boldsymbol{\theta} - \theta^\circ\|]^{1/2}$. Define the tight sequence of random variables $S_n := (\|\boldsymbol{\theta} - \theta^\circ\| - E)/V$. We consider the sequence of events $\Omega_n := \{E \geqslant c_n\Phi_n\} \cup \{V \geqslant c_n\Phi_n\}$. We have $\mathbb{P}(\Omega_n) \leqslant \mathbb{P}(\{E \geqslant c_n\Phi_n\}) + \mathbb{P}(\{V \geqslant c_n\Phi_n\})$ which hence tends to 0. Hence, for $K_n := c_n\Phi_n(1 + c_n)$, we can write

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(\|\boldsymbol{\theta} - \theta^\circ\| > K_n)\right] = \mathbb{E}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(S_n > (K_n - E)/V)\right] \\
&\quad = \mathbb{E}\left[\mathbb{1}_{\Omega_n}\mathbb{P}_{\boldsymbol{\theta}|Y^n}(S_n > (K_n - E)/V)\right] + \mathbb{E}\left[\mathbb{1}_{\Omega_n^c}\mathbb{P}_{\boldsymbol{\theta}|Y^n}(S_n > (K_n - E)/V)\right] \\
&\quad \leqslant \mathbb{P}(\Omega_n) + \mathbb{P}(\Omega_n^c)\cdot\mathbb{E}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}\left(S_n > \frac{K_n - c_n\Phi_n}{c_n\Phi_n}\right)\right] \leqslant \mathbb{P}(\Omega_n) + \mathbb{E}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}(S_n > c_n)\right].
\end{aligned}$$

We can conclude as S_n is a tight sequence, c_n tends to infinity and $\mathbb{P}(\Omega_n)$ tends to 0. □

2.3.2 Employing exponential concentration inequalities

We give here the structure of the proof we use to prove the optimality of the (finitely) iterated hierarchical sieve prior. This method takes advantage of the structure of the hierarchical prior and the specific form of the l^2 norm. It is similar to the one used in Johannes et al. (2014).
Its main argument is an interpretable decomposition of the risk. Under the assumption that those parts can be controlled properly (which has to be checked using properties of the considered model), one obtains a contraction rate.
Let us first present the set of assumptions which has to be verified depending on the considered model.

Assumption 12 Assume that one can find three sequences $(G_n^-)_{n\in\mathbb{N}}$, (m_n°) and $(G_n^+)_{n\in\mathbb{N}}$ in $\mathbb{M}$ such that, for any n, we have $0 \leqslant G_n^- \leqslant m_n^\circ \leqslant G_n^+ \leqslant G_n$; two sequences of real numbers $(K_{A,n})_{n\in\mathbb{N}}$ and $(K_{B,n})_{n\in\mathbb{N}}$ such that the following properties hold.

First, we assume that values of M which are "too small" have a small probability. The sequence of events $\mathcal{A}_{m,n} := \{\Upsilon^\eta(Y^n, m_n^\circ) - \Upsilon^\eta(Y^n, m) < K_{A,n}\}$ verifies

$$\sum\nolimits_{m<G_n^-} \exp\left[-\eta\left(K_{A,n} + (\mathrm{pen}(m_n^\circ) - \mathrm{pen}(m))\right)\right] \in \mathfrak{o}_n(1); \quad \sum\nolimits_{m<G_n^-} \mathbb{P}\left[\mathcal{A}_{m,n}^c\right] \in \mathfrak{o}_n(1)$$

Secondly, we assume that M takes large values with small probability. That is to say, the sequence of events $\mathcal{B}_{m,n} := \{\Upsilon^\eta(Y^n, m) - \Upsilon^\eta(Y^n, m_n^\circ) < K_{B,n}\}$ verifies

$$\sum\nolimits_{m>G_n^+} \exp\left[-\eta\left(K_{B,n} + (\mathrm{pen}(m_n^\circ) - \mathrm{pen}(m))\right)\right] \in \mathfrak{o}_n(1); \quad \sum\nolimits_{m<G_n^-} \mathbb{P}\left[\mathcal{B}_{m}^c\right] \in \mathfrak{o}_n(1).$$

Finally we assume that, if M lends between G_n^- and G_n^+, the posterior behaves properly.

$$\sum\nolimits_{G_n^- \leqslant m \leqslant G_n^+} \mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{m}} - \theta^\circ\|_{l^2}^2 > \Phi_n\right)\right] \in \mathfrak{o}_n(1).$$

□

Under this set of hypotheses one obtains that Φ_n is a contraction rate for the posterior distribution.

Theorem 2.3.2.
Under Assumption 12, for any η in $[\![1, \infty[\![$ there exists a constant K such that

$$\lim\nolimits_{n\to\infty} \mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 \geqslant K\Phi_n\right)\right] = 0.$$

□

Proof of Theorem 2.3.2
First, notice that we have the following decomposition:

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 > \Phi_n\right)\right] \\
&\quad = \sum\nolimits_{m\leqslant G_n} \mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left(\left\{\|\boldsymbol{\theta}_{\overline{m}} - \theta^\circ\|_{l^2}^2 > \Phi_n\right\}\right) \cdot \mathbb{P}^{(\eta)}_{M|Y^n}\left(\{m\}\right)\right].
\end{aligned}$$

Then, for any three sequences m_n°, G_n^+ and G_n^- with $G_n^- \leqslant m_n^\circ \leqslant G_n^+ \leqslant G_n$, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 > \Phi_n\right)\right] \\
&\quad \leqslant \underbrace{\mathbb{E}\left[\mathbb{P}^{(\eta)}_{M|Y^n}\left(M < G_n^-\right)\right]}_{=:A} + \underbrace{\mathbb{E}\left[\mathbb{P}^{(\eta)}_{M|Y^n}\left(M > G_n^+\right)\right]}_{=:B} \\
&\quad + \sum\nolimits_{G_n^- \leqslant m \leqslant G_n^+} \underbrace{\mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left(\left\{\|\boldsymbol{\theta}_{\overline{m}} - \theta^\circ\|_{l^2}^2 > \Phi_n\right\}\right)\right]}_{=:C_m}.
\end{aligned}$$

The goal is then to control the three sums using concentration inequalities.
We begin with A, where the conclusion is given by Assumption 12:

$$\begin{aligned}A = &\sum_{m<G_n^-} \mathbb{E}\left[\exp\left[\eta\left(\mathrm{pen}(m) + \Upsilon^\eta(Y^n, m)\right)\right] / \sum_{j\leqslant G} \exp\left[\eta\left(\mathrm{pen}(j) + \Upsilon^\eta(Y^n, j)\right)\right] \mathbb{1}_{\mathcal{A}_m}\right]\\ &+ \mathbb{E}\left[(\exp\left[\eta\left(\mathrm{pen}(m) + \Upsilon^\eta(Y^n, m)\right)\right])/(\textstyle\sum_{j\leqslant G} \exp\left[\eta\left(\mathrm{pen}(j) + \Upsilon^\eta(Y^n, j)\right)\right])\mathbb{1}_{\mathcal{A}_m^c}\right]\\ &\leqslant \textstyle\sum_{m<G_n^-} \exp\left[-\eta\left(K_{A,n} + \left(\mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)\right)\right)\right] + \mathbb{P}\left[\mathcal{A}_m^c\right] \in \mathfrak{o}_n(1).\end{aligned}$$

We process similarly for B, where the conclusion is given by Assumption 12:

$$\begin{aligned}B = &\sum_{m>G_n^+} \mathbb{E}\left[(\exp\left[\eta\left(\mathrm{pen}(m) + \Upsilon^\eta(Y^n, m)\right)\right])/(\sum_{j\leqslant G_n} \exp\left[\eta\left(\mathrm{pen}(j) + \Upsilon^\eta(Y^n, j)\right)\right])\mathbb{1}_{\mathcal{B}_m}\right]\\ &+ \mathbb{E}\left[(\exp\left[\eta\left(\mathrm{pen}(m) + \Upsilon^\eta(Y^n, m)\right)\right])/(\textstyle\sum_{j\leqslant G_n} \exp\left[\eta\left(\mathrm{pen}(j) + \Upsilon^\eta(Y^n, j)\right)\right])\mathbb{1}_{\mathcal{B}_m^c}\right]\\ &\leqslant \textstyle\sum_{m>G_n^+} \exp\left[-\eta\left(K_{B,n} + \left(\mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)\right)\right)\right] + \mathbb{P}\left[\mathcal{B}_m^c\right] \in \mathfrak{o}_n(1)\end{aligned}$$

Finally, C_m is directly controlled by Assumption 12.

□

2.3.3 Generalisation for self informative Bayes carrier

In the previous section, we described the kind of proof used in Johannes et al. (2014) and argued that it can also be used with a finitely iterated posterior. We present here an adaptation of this scheme for the self informative Bayes carrier. The main subtlety lies in the fact that the hyper-parameter only loads extrema of a penalised contrast function.
We first adapt the set of assumptions.

Assumption 13 Assume that one can find three sequences $(G_n^-)_{n\in\mathbb{N}}$, (m_n°) and $(G_n^+)_{n\in\mathbb{N}}$ in $\mathbb{M}$ such that, for any n, we have $0 \leqslant G_n^- \leqslant m_n^\circ \leqslant G_n^+ \leqslant G_n$ such that the following properties hold true:

$$\begin{aligned}&\textstyle\sum_{m<G_n^-} \mathbb{P}\left(\Upsilon(m, Y^n) - \Upsilon(m_n^\circ, Y^n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)\right) &&\in \mathfrak{o}_n(1);\\ &\textstyle\sum_{m>G_n^+} \mathbb{P}\left(\Upsilon(m, Y^n) - \Upsilon(m_n^\circ, Y^n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)\right) &&\in \mathfrak{o}_n(1);\\ &\textstyle\sum_{G_n^-\leqslant m\leqslant G_n^+} \mathbb{P}\left[\|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2 > \Phi_n\right] &&\in \mathfrak{o}_n(1).\end{aligned}$$

□

Those assumptions can generally be obtained in a similar fashion as those in Assumption 12. We then obtain a similar result for the contraction of the self informative Bayes carrier.

Theorem 2.3.3.
Under Assumption 13, there exists a constant K such that

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|^2_{l^2} \geqslant K\Phi_n\right)\right] = 0.$$

□

Proof of Theorem 2.3.3
We start the proof in a similar fashion to Theorem 2.3.2:

$$\begin{aligned}&\mathbb{E}\left[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|^2_{l^2} > \Phi_n\right)\right]\\ &= \sum_{m\leqslant G_n} \mathbb{E}\left[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left(\left\{\|\boldsymbol{\theta}_{\overline{m}} - \theta^\circ\|^2_{l^2} > \Phi_n\right\}\right)\cdot \mathbb{P}^{(\infty)}_{M|Y^n}(\{m\})\right].\end{aligned}$$

Then, for any three subsets m_n°, G_n^+ and G_n^- with $0 \leqslant G_n^- \leqslant m_n^\circ \leqslant G_n^+ \leqslant G_n$, we have

$$\begin{aligned}&\mathbb{E}\left[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|^2_{l^2} > \Phi_n\right)\right]\\ &\leqslant \underbrace{\mathbb{E}\left[\mathbb{P}^{(\eta)}_{M|Y^n}(M < G_n^-)\right]}_{=:A} + \underbrace{\mathbb{E}\left[\mathbb{P}^{(\eta)}_{M|Y^n}(M > G_n^+)\right]}_{=:B}\\ &+ \underbrace{\sum_{G_n^-\leqslant m\leqslant G_n^+} \mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left(\left\{\|\boldsymbol{\theta}_{\overline{m}} - \theta^\circ\|^2_{l^2} > \Phi_n\right\}\right)\right]}_{=:C_m}\end{aligned}$$

The goal is then to control the three sums using concentration inequalities.
We begin with A, the conclusion is given by Assumption 13:

$$\begin{aligned}A &= \mathbb{P}\left[\forall l \geqslant G_n^-, \mathrm{pen}(\widehat{m}) + \Upsilon(\widehat{m}, Y^n) < \mathrm{pen}(l) + \Upsilon(l, Y^n)\right]\\ &\leqslant \mathbb{P}\left[\exists m < G_n^-, \mathrm{pen}(m) + \Upsilon(m, Y^n) < \mathrm{pen}(m_n^\circ) + \Upsilon(m_n^\circ, Y)\right]\\ &\leqslant \sum_{m<G_n^-} \mathbb{P}\left[\mathrm{pen}(m) + \Upsilon(m, Y^n) < \mathrm{pen}(m_n^\circ) + \Upsilon(m_n^\circ, Y^n)\right]\\ &\leqslant \sum_{m<G_n^-} \mathbb{P}\left[\Upsilon(m, Y^n) - \Upsilon(m_n^\circ, Y^n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)\right] \in \mathfrak{o}_n(1).\end{aligned}$$

We process similarly for B, the conclusion is given by Assumption 13:

$$\begin{aligned}B &= \mathbb{P}\left[\forall l \leqslant G_n^+, \mathrm{pen}(\widehat{m}) + \Upsilon(\widehat{m}, Y^n) < \mathrm{pen}(l) + \Upsilon(l, Y^n)\right]\\ &\leqslant \mathbb{P}\left[\exists m > G_n^+, \mathrm{pen}(m) + \Upsilon(m, Y^n) < \mathrm{pen}(m_n^\circ) + \Upsilon(m_n^\circ, Y^n)\right]\\ &\leqslant \sum_{m>G_n^+} \mathbb{P}\left[\mathrm{pen}(m) + \Upsilon(m, Y^n) < \mathrm{pen}(m_n^\circ) + \Upsilon(m_n^\circ, Y^n)\right]\\ &\leqslant \sum_{m>G_n^+} \mathbb{P}\left[\Upsilon(m, Y^n) - \Upsilon(m_n^\circ, Y^n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)\right] \in \mathfrak{o}_n(1).\end{aligned}$$

Finally, C_m is directly controlled by Assumption 13.

□

2.4 Application to the inverse Gaussian sequence space model

In this section, we consider the inverse Gaussian sequence space model as introduced in Definition 33. First, we investigate about the self informative Bayes limit/carrier of (hierarchical) Gaussian sieve priors using the technics presented in Theorem 2.1.1, Theorem 2.2.1 and Theorem 2.2.2. Then, we use the methodology described in section 2.3.1 to compute upper bounds of the Gaussian sieve priors described in section 2.1 when applied to this specific model. Doing so, we will notice that it gives us, for a general case, the same speed as the convergence rate of projection estimators and that, by choosing properly the threshold parameter, we reach the oracle rate of convergence as well as the minimax optimal rate, **without a log-loss**. However, we also see that this strategy cannot be applied to the hierarchical priors we are interested in. Hence, we then use the strategy exposed in section 2.3.2 and show that under some regularity conditions, the iterated hierarchical prior leads to optimal posterior contraction rate. As a consequence, we can conclude about the oracle and minimax speed of convergence of the penalised contrast model selection estimator with a new strategy of proof.

2.4.1 Self informative Bayes carrier for Gaussian sieve in iGSSM

We first consider the asymptotic $\eta \to \infty$ for the Gaussian sieve prior.

THEOREM 2.4.1.
For a Gaussian sieve prior with threshold parameter m, the self informative Bayes carrier is the singleton given by: $\theta_{n,\overline{m}} = (\theta_{n,\overline{m}}(s))_{s\in\mathbb{N}} = \left(\phi_n(s)\lambda^{-1}(s)\mathbb{1}_{|s|\leqslant m}\right)_{s\in\mathbb{N}}$. □

PROOF OF THEOREM 2.4.1
In this model, we explicitly have that $\mathbb{M} = \mathbb{N}$; in addition, for any θ in $\Theta_{\overline{m}}$, and ϕ_n in Θ, there exists C only depending on ϕ_n and n such that,

$$l(\theta, \phi_n) = -n^{-1/2}\left(\sum\nolimits_{s\leqslant m} \phi_n(s)\lambda(s)\theta(s) - \sum\nolimits_{s\leqslant m} \Lambda(s)^{-1}\theta(s)^2/2\right) + C;$$

which is continuous with respect to θ; therefore, Assumption 10 is verified.
We can hence apply Theorem 2.1.1 which proves that the support of the self informative Bayes carrier is contained in the set of maximisers of $l(\theta, \phi_n)$ which is obviously the singleton $\{\left(\theta_n(s)\lambda^{-1}(s)\mathbb{1}_{|s|\leqslant m}\right)_{s\in\mathbb{N}}\}$. □

As an alternative, one could have noticed that the prior and likelihood are conjugated. Define for any s in $\mathbb{N}$ and η in $\mathbb{N}^\star$ the quantities

$$\widetilde{\theta}^{(\eta)}(s) := (n\eta\phi_n(s)\lambda(s))/(1 + n\eta\lambda(s)^2); \quad \sigma^{(\eta)}(s) := (1 + n\eta\lambda(s)^2)^{-1}.$$

Then, for any s in $\mathbb{N}$, the posterior distribution of $\boldsymbol{\theta}(s)$ after η iterations is given by

$$\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}(s)|\phi_n} = \mathcal{N}(\widetilde{\theta}^{(\eta)}(s), \sigma^{(\eta)}(s))\mathbb{1}_{|s|\leqslant m} + \delta_0(\boldsymbol{\theta}(s))\mathbb{1}_{|s|>m}.$$

Considering the respective limits of $\widetilde{\theta}^{(\eta)}(s)$ and $\sigma^{(\eta)}(s)$ as η tends to ∞ for any s in $\mathbb{N}$ coincides with our previous statement.

2.4.2 Contraction rate for Gaussian sieve in iGSSM

We now investigate the behaviour of the Gaussian sieve prior applied to iGSSM as n tends to ∞. In this context, it is interesting to let η and m depend on n; we hence note η_n and m_n.

First consider the strategy exposed in section 2.3.1. To apply it, we will place ourselves under the following hypothesis that apparently limits the possible choices for the threshold parameter. In practice, the thresholds which are left aside would be too large and are known to lead to a poor estimation performance.

Assumption 14 Assume that m_n and η_n are chosen in such a way that either of

- $\sum_{s\leqslant m_n} \Lambda(s)n^{-1} \in \mathcal{O}_n(1)$
- $\sum_{s\leqslant m_n} (\Lambda(s)|\theta^{\circ}(s)|)^2(n\eta_n)^{-2} \in \mathcal{O}_n\left(\sum_{s\leqslant m_n} \Lambda(s)n^{-1}\right)$ and $\sum_{s\leqslant m_n} (\Lambda(s)^{3/2}|\theta^{\circ}(s)|)(n^{3/2}\eta_n)^{-1} \in \mathcal{O}_n\left(\sum_{s\leqslant m_n} \Lambda(s)n^{-1}\right)$

stand true. □

We illustrate this hypothesis under the typical behaviours of θ and λ

Numerical discussion 2.4.1.
Consider the first inclusion $\sum_{s\leqslant m_n} \Lambda(s)/n \in \mathcal{O}_n(1)$.
Notice that (p) and (np) have no influence here.

[p-o], **[o-o]**, and **[s-o]** we have $\sum_{s\leqslant m_n} \Lambda(s)/n = n^{-1}m\Lambda_{\circ}(m_n) \approx n^{-1}m_n^{2a+1}$ and hence the first inclusion is equivalent to $m_n \in \mathcal{O}_n(n^{1/(2a+1)})$.

[p-s], and **[o-s]** we have $\sum_{s\leqslant m_n} \Lambda(s)/n = n^{-1}m\Lambda_{\circ}(m_n) \approx n^{-1}m_n^{-(1-2a)_+}\exp[m_n^{2a}]$ and hence the first inclusion is equivalent to $m_n \in \mathcal{O}_n(\log(n)^{1/(2a)})$.

In the second inclusion, $\sum_{s\leqslant m_n} (\Lambda(s)|\theta^{\circ}(s)|)^2/(n\eta_n)^2 \in \mathcal{O}\left(\sum_{s\leqslant m_n} \Lambda(s)/n\right)$, the regularity of θ also intervenes. Notice that, under **[o-o]** and **[o-s]**, $\sum_{s\leqslant m_n} \Lambda(s)/n \approx n^{-1}m_n^{2a+1}$ while under **[s-o]** we have $\sum_{s\leqslant m_n} \Lambda(s)/n \approx n^{-1}m_n^{-(1-2a)_+}\exp[m_n^{2a}]$.

(p) $\sum_{s\leqslant m_n} (\Lambda(s)|\theta^{\circ}(s)|)^2/(n\eta_n)^2 \leqslant \sum_{s\leqslant K} (\Lambda(s)|\theta^{\circ}(s)|)^2/(n\eta_n)^2 \in \mathfrak{o}_n(n^{-1})$ and hence the inclusion is always verified.

(np) We now have to distinguish the different regularities of θ and λ. In any case, notice that $\sum_{s\leqslant m_n} (\Lambda(s)|\theta^{\circ}(s)|)^2/(n\eta_n)^2 \leqslant (n\eta_n)^{-2}\sum_{s\leqslant m_n} \Lambda(s)^2 \cdot (\|\theta^{\circ}\|_{l^2}^2 - \mathfrak{b}_m^2(\theta^{\circ}))$

[o-o] $(n\eta_n)^{-2} \sum\limits_{s\leqslant m_n} \Lambda(s)^2 \cdot \sum\limits_{s\leqslant m_n} |\theta^{\circ}(s)|^2 \approx (n\eta_n)^{-2} \cdot m^{4a+1}$ implies $m_n \in \mathcal{O}_n(n^{1/(2a)}\eta_n^{1/a})$;

[o-s] $(n\eta_n)^{-2} \sum\limits_{s\leqslant m_n} \Lambda(s)^2 \cdot \sum\limits_{s\leqslant m_n} |\theta^{\circ}(s)|^2 \approx (n\eta_n)^{-2}m^{-(1-4a)_+}\exp[m^{4a}]$ implies $m_n \in \mathcal{O}_n(\log(n\eta_n^2)^{1/(4a)})$;

[s-o] $(n\eta_n)^{-2} \sum\limits_{s\leqslant m_n} \Lambda(s)^2 \cdot \sum\limits_{s\leqslant m_n} |\theta^{\circ}(s)|^2 \approx (n\eta_n)^{-2} \cdot m^{4a+1}$ implies $m_n \in \mathcal{O}_n(n^{1/(2a)}\eta_n^{1/a})$.

Finally, for the third inclusion $\sum_{s\leqslant m_n}(\Lambda(s)^{3/2}|\theta^\circ(s)|)/(n^{3/2}\eta_n) \in \mathcal{O}\left(\sum_{s\leqslant m_n}\Lambda(s)/n\right)$ notice that we have $\sum_{s\leqslant m_n}(\Lambda(s)^{3/2}|\theta^\circ(s)|)/(n^{3/2}\eta_n) \leqslant (n^{3/2}\eta_n)^{-1}\cdot\sum_{s\leqslant m_n}\Lambda(s)^3\cdot(\|\theta^\circ\|_{l^2}^2 - \mathfrak{b}_{m_n}^2(\theta^\circ))$. Under **[o-o]** and **[o-s]**, $\sum_{s\leqslant m_n}\Lambda(s)/n \approx n^{-1}m_n^{2a+1}$ while under **[s-o]** we have $\sum_{s\leqslant m_n}\Lambda(s)/n \approx n^{-1}m_n^{-(1-2a)_+}\exp[m_n^{2a}]$.

(p) $\sum_{s\leqslant m_n}(\Lambda(s)^{3/2}|\theta^\circ(s)|)/(n^{3/2}\eta_n) \leqslant (n^{3/2}\eta_n)^{-1}\sum_{s\leqslant K}(\Lambda(s)^{3/2}|\theta^\circ(s)|) \in \mathfrak{o}_n(n^{-1})$ and hence the inclusion is always verified.

(np) We now have to distinguish the different regularities of θ and λ.

- **[o-o]** $(n^{3/2}\eta_n)^{-1}\cdot\sum_{s\leqslant m_n}\Lambda(s)^3\cdot(\|\theta^\circ\|_{l^2}^2 - \mathfrak{b}_{m_n}^2(\theta^\circ)) \approx (n^{3/2}\eta_n)^{-1}\cdot m^{6a+1}$ implies $m_n \in \mathcal{O}_n((\eta_n\sqrt{n})^{1/(4a)})$;
- **[o-s]** $(n^{3/2}\eta_n)^{-1}\cdot\sum_{s\leqslant m_n}\Lambda(s)^3\cdot(\|\theta^\circ\|_{l^2}^2 - \mathfrak{b}_{m_n}^2(\theta^\circ)) \approx (n^{3/2}\eta_n)^{-1}\cdot m^{-(1-6a)_+}\exp[m^{6a}]$ implies $m_n \in \mathcal{O}_n(\log(\sqrt{n}\eta_n)11/(6a))$;
- **[s-o]** $(n^{3/2}\eta_n)^{-1}\cdot\sum_{s\leqslant m_n}\Lambda(s)^3\cdot(\|\theta^\circ\|_{l^2}^2 - \mathfrak{b}_{m_n}^2(\theta^\circ)) \approx (n^{3/2}\eta_n)^{-1}\cdot m^{6a+1}$ implies $m_n \in \mathcal{O}_n((\eta_n\sqrt{n})^{1/(4a)})$.

□

We see that in any case, one can chose the sequence $(\eta_n)_{n\in\mathbb{N}}$ in such a way that the condition is weaker that $m_n \in \mathcal{O}_n(n)$; unfortunately, this choice generally depends on the ill-posedness parameter a and adaptively chosing η is not considered here.
Under this hypothesis we can obtain the contraction rate we hoped for.

Corollary 2.4.1.
Under Assumption 14, for any θ° in Θ and increasing, unbounded sequence c_n, we have

$$\lim_{n\to\infty}\mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left(\|\theta^\circ - \boldsymbol{\theta}_{\overline{m_n}}\|_{l^2}^2 \leqslant c_n\Phi_n^{m_n}\right)\right] = 1.$$

□

Proof of corollary 2.4.1
Remind that, for any s in $\mathbb{N}$, $\phi_n(s) = \phi(s) + n^{-1/2}\xi(s)$, where $(\xi(s))_{s\in\mathbb{N}}$ is an iid. Gaussian white noise process. We will apply Theorem 2.3.1 and hence need to show:

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_{\boldsymbol{\theta}|\phi_n}\left[\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2\right]\right] \in \mathcal{O}_n(\Phi_n^{m_n}); \quad \mathbb{V}\left[\mathbb{E}_{\boldsymbol{\theta}|\phi_n}\left[\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2\right]\right]^{1/2} \in \mathcal{O}_n(\Phi_n^{m_n});\\
&\mathbb{E}\left[\mathbb{V}_{\boldsymbol{\theta}|\phi_n}\left[\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2\right]^{1/2}\right] \in \mathcal{O}_n(\Phi_n^{m_n}); \quad \mathbb{V}\left[\mathbb{V}_{\boldsymbol{\theta}|\phi_n}\left[\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2\right]^{1/2}\right]^{1/2} \in \mathcal{O}_n(\Phi_n^{m_n}).
\end{aligned}$$

We use the fact that $\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2 = \sum_{|s|\leqslant m_n}(\boldsymbol{\theta}(s) - \theta^\circ(s))^2 + \mathfrak{b}_{m_n}^2(\theta^\circ)$ and that we know the distribution of $\boldsymbol{\theta}(s)$. This gives us the expectation and variance of the posterior distribution of $\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2$. We use in addition $(1 + \Lambda(s)/(n\eta_n))^{-1} \leqslant 1$ to obtain upper bounds for these quantities.

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta} - \theta^\circ\|_{l^2}^2\right] &= \sum_{|s|\leqslant m_n}\left(\frac{\Lambda(s)/(n\eta_n)}{\Lambda(s)/(n\eta_n)+1}\right)\left(1 + \frac{\left(-\theta^\circ(s) + \eta_n\sqrt{n}\xi(s)\lambda(s)\right)^2}{(\eta_n n)/\Lambda(s)+1}\right) + \mathfrak{b}_{m_n}^2\\
&\leqslant \sum_{|s|\leqslant m_n}(\Lambda(s)/n\eta_n) + \sum_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^2\left(-\theta^\circ(s) + \eta_n\sqrt{n}\xi(s)\lambda(s)\right)^2 + \mathfrak{b}_{m_n}^2(\theta^\circ);
\end{aligned}$$

$$\mathbb{V}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}-\theta^\circ\|_{l^2}^2\right] = 2\sum_{|s|\leqslant m_n}\left(\frac{\Lambda(s)/(n\eta_n)}{\Lambda(s)/(n\eta_n)+1}\right)^2\left(1+2\frac{(-\theta^\circ(s)+\eta_n\sqrt{n}\xi(s)\lambda(s))^2}{(\eta_n n)/\Lambda(s)+1}\right)$$
$$\leqslant 2\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^2 + 4\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^3\left(-\theta^\circ(s)+\eta_n\sqrt{n}\xi(s)\lambda(s)\right)^2.$$

In addition, we use the sub-additivity of the square root to obtain this upper bound:

$$\mathbb{V}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}-\theta^\circ\|_{l^2}^2\right]^{1/2}$$
$$\leqslant \sqrt{2}\sum\nolimits_{|s|\leqslant m_n}\Lambda(s)/(n\eta_n) + 2\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^{3/2}\left|-\theta^\circ(s)+\eta_n\sqrt{n}\xi(s)\lambda(s)\right|.$$

Using linearity of the expectation and the standard Gaussian distribution of ξ_j we have:

$$\mathbb{E}\left[\mathbb{E}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}-\theta^\circ\|_{l^2}^2\right]\right]$$
$$\leqslant \sum\nolimits_{|s|\leqslant m_n}\Lambda(s)/(n\eta_n) + \sum\nolimits_{|s|\leqslant m_n}\Lambda(s)/n + \sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^2\left|\theta^\circ(s)\right|^2 + \mathfrak{b}_{m_n}^2(\theta^\circ).$$

The same properties give us this bound:

$$\mathbb{V}\left[\mathbb{E}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}-\theta^\circ\|_{l^2}^2\right]\right] \leqslant 2\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/n)^2 + 4\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)^3/(\eta_n^2 n^3))\left|\theta^\circ(s)\right|^2;$$

and we use the sub-additivity of the square root in addition:

$$\mathbb{V}\left[\mathbb{E}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}-\theta^\circ\|_{l^2}^2\right]\right]^{1/2} \leqslant \sqrt{2}\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/n) + 2\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)^{3/2}/(\eta_n n^{3/2}))\left|\theta^\circ(s)\right|.$$

To control the moments of the posterior variance, we use the properties of the folded Gaussian random variables:

$$\mathbb{E}\left[\mathbb{V}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}-\theta^\circ\|_{l^2}^2\right]^{1/2}\right]$$
$$\leqslant \sqrt{2}\sum_{|s|\leqslant m_n}\Lambda(s)/(n\eta_n) + 2\sum_{|s|\leqslant m_n}(2/(\pi\cdot n^3\eta_n))^{1/2}\Lambda(s)\exp\left[-((\theta^\circ(s))^2\Lambda(s))/(2\eta_n^2)\right]$$
$$+\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^{3/2}|\theta^\circ(s)|;$$

$$\mathbb{V}\left[\mathbb{V}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}_{\overline{m_n}}-\theta^\circ\|_{l^2}^2\right]^{1/2}\right] \leqslant 2\sum\nolimits_{|s|\leqslant m_n}(\Lambda(s)/(n\eta_n))^3\cdot\left[|\theta^\circ(s)|^2+\eta_n^2/\Lambda(s)\right];$$

$$\mathbb{V}\left[\mathbb{V}_{\boldsymbol{\theta}_{\overline{m_n}}|\phi_n}\left[\|\boldsymbol{\theta}_{\overline{m_n}}-\theta^\circ\|_{l^2}^2\right]^{1/2}\right]^{1/2}$$
$$\leqslant \sqrt{2}\sum\nolimits_{|s|\leqslant m_n}((\Lambda(s)^3\left|\theta^\circ(s)\right|^2)(n\eta_n)^{-3})^2 + \sum\nolimits_{|s|\leqslant m_n}\Lambda(s)/(n^3\eta_n)^{1/2}.$$

Using Assumption 14, the leading term in each of these bounds is for the most of order $\Phi_n^{m_n}$ and hence, we can apply Theorem 2.3.1 which proves the statement. □

Notice that if one selects $m_n = m_n^\circ$ we obtain the oracle rate of convergence of projection estimators.

COROLLARY 2.4.2.
For any θ° in Θ and increasing, unbounded sequence c_n, we have

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m_n^\circ}}|\phi_n}\left(\left\|\theta^\circ - \boldsymbol{\theta}_{\overline{m_n^\circ}}\right\|^2_{l^2} \leqslant c_n \Phi^\circ_n\right)\right] = 1.$$

□

We have hence seen that Gaussian sieve priors contract around the true parameter at the same rate as the projection estimator with identical threshold parameter contract and that, in particular, the best Gaussian sieve prior contracts at the oracle convergence rate of the projection estimators.

2.4.3 Self informative Bayes carrier for the hierarchical prior

In this subsection, we propose an analytical shape for a hierarchical Gaussian sieve prior to use in the context of an inverse Gaussian sequence space model.
We doubly justify this choice, first by showing that the self informative limit is a penalised contrast maximiser projection estimator and, in the next subsection, that this choice yields good contraction properties.

First remind that for any s in $\mathbb{N}$, we have:

$$\widetilde{\theta}^{(\eta)}(s) = (n\eta\phi_n(s)\lambda(s)) \cdot (1 + n\eta\lambda(s)^2)^{-1}; \text{ and } \quad \sigma^{(\eta)}(s) = (1 + n\eta|\lambda(s)|^2)^{-1};$$

and define for any m in $\mathbb{N}$ the notations

$$\sigma^{(\eta)}_{\overline{m}} := (\sigma^{(\eta)}(s)\mathbb{1}_{\{s\leqslant m\}})_{s\in\mathbb{N}}; \text{ and } \quad \widetilde{\theta}^{(\eta)}_{\overline{m}} := (\widetilde{\theta}^{(\eta)}(s)\mathbb{1}_{\{s\leqslant m\}})_{s\in\mathbb{N}}.$$

Then, we define, for any m in $\mathbb{N}$, the quantity $\Lambda_+(m) := \max_{s\leqslant m}\{\Lambda(s)\}$. We then take $G_n := \max\{m \in [\![1, n]\!] : \Lambda_+(m)/n \leqslant \Lambda(0)\}$.
For any m in $\mathbb{N}$, we make the following choice for the prior distribution of M

$$\mathbb{P}_M(\{m\}) \propto \exp\left[-\eta/2\left(3m + \sum\nolimits_{s=0}^{m} \log(\sigma^\eta(s))\right)\right].$$

Using the notations of section 2.2 (and keeping in mind the notation for weighted norms given in section 1.3.2.1 in the context of Sobolev's ellipsoid, and the convention "$0/0 = 0$"), we have

$$\begin{aligned}
\mathrm{pen}(m) \ &= (\eta/2)\left(3m + \sum\nolimits_{s=0}^{m} \log(\sigma^{(\eta)}(s))\right); \\
\Upsilon^\eta(Y, m) &= \sum\nolimits_{s=0}^{m} n|\phi_n(s)|^2\left(\Lambda(s)(n\eta)^{-1} + 1\right)^{-1} + (1/2)\sum\nolimits_{s=0}^{m} \log(\sigma^{(\eta)}(s)).
\end{aligned}$$

Which leads us to the iterated prior of the hyper-parameter:

$$\mathbb{P}^{(\eta)}_{M|\phi_n}(m) \propto \exp\Big[-(\eta/2)\left(3m - n\sum\nolimits_{s\leqslant m} |\phi_n(s)|^2(\Lambda(s)(n\eta)^{-1} + 1)^{-1}\right)\Big].$$

We can hence simplify our notation in the following way: $\mathrm{pen}(m) = 3m$ and $\Upsilon^\eta(Y, m) = \sum_{s\leqslant m} n|\phi_n(s)|^2(\Lambda(s)(n\eta)^{-1} + 1)^{-1}$. Let us remind that the iterated distribution for $\boldsymbol{\theta}_{\overline{M}}|\phi_n$

is given by $\mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n} = \sum_{m\in\mathbb{N}} \mathbb{P}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|\phi_n} \cdot \mathbb{P}^{(\eta)}_{M|\phi_n}(m)$. Hence, according to Theorem 2.2.2, the self informative limit for the hyper-parameter is $\widehat{m} := \arg\min_{m\leqslant G_n}\{3m - n\sum_{s\leqslant m}|\phi_n(s)|^2\}$; and the self informative Bayes limit for $\boldsymbol{\theta}_{\overline{M}}$ is the associated projection estimator $\theta_{n,\overline{m}}$.

Note that, defining for any m in $[\![1, G_n]\!]$ the quantity $E(m) = 3m - n\sum_{s\leqslant m}|\phi_n(s)|^2$; for all distinct k and m in $[\![1, G_n]\!]$, we almost surely have $E(k) - E(m) \neq 0$ since $\Upsilon(k) - \Upsilon(m)$ is a random variable with absolutely continuous distribution with respect to Lebesgue measure and hence, $\mathbb{P}_{\theta^\circ}[\{\Upsilon(k) - \Upsilon(m) = \mathrm{pen}(k) - \mathrm{pen}(m)\}] = 0$.

2.4.4 Contraction rate for the hierarchical prior

In this subsection, we discuss the contraction rate of the hierarchical Gaussian iterated posterior distribution by applying the methodology described in section 2.3.2.
The results are similar to the ones obtained in Johannes et al. (2014) but extended to the iterated posterior distribution, included in the case of "$\eta = \infty$", in such a way that it offers a novel proof for optimality of the penalised contrast maximiser projection estimator.
Remind that we defined for any m in $\mathbb{N}$ the quantities $\Lambda_+(m) = \max_{|s|\leqslant m}\{\Lambda(s)\}$ and $\Lambda_\circ(m) = m^{-1}\sum_{|s|\leqslant m}\Lambda(s)$.
The results are obtained using the following contraction inequalities, which can be found in this form in Johannes et al. (2014) as a result adapted from Birgé (2001) and Laurent et al. (2012).

Lemma 2.4.1.
Let $\{X(s)\}_{s\geqslant 1}$ be independent and normally distributed random variables with real mean $\alpha(s)$ and standard deviation $\beta(s) \geqslant 0$. For $m \in \mathbb{N}$, set $S_m := \sum_{s=1}^{m} X(s)^2$ and consider $v_m \geqslant \sum_{s=1}^{m}\beta(s)^2$, $t_m \geqslant \max_{1\leqslant s\leqslant m}\beta(s)^2$ and $r_m \geqslant \sum_{s=1}^{m}\alpha(s)^2$. Then for all $c \geqslant 0$, we have

$$\sup_{m\geqslant 1}\exp\left[\frac{c(c\wedge 1)(v_m + 2r_m)}{4t_m}\right]\mathbb{P}\left(S_m - \mathbb{E}[S_m] \leqslant -c(v_m + 2r_m)\right) \leqslant 1;$$
$$\sup_{m\geqslant 1}\exp\left[\frac{c(c\wedge 1)(v_m + 2r_m)}{4t_m}\right]\mathbb{P}\left(S_m - \mathbb{E}[S_m] \geqslant \frac{3c}{2}(v_m + 2r_m)\right) \leqslant 1.$$

□

Lemma 2.4.2.
Let $\{X(s)\}_{s\geqslant 1}$ be independent and normally distributed random variables with real mean $\alpha(s)$ and standard deviation $\beta(s) \geqslant 0$. For $m \in \mathbb{N}$, set $S_m := \sum_{s=1}^{m} X(s)^2$ and consider $v_m \geqslant \sum_{s=1}^{m}\beta(s)^2$, $t_m \geqslant \max_{1\leqslant s\leqslant m}\beta(s)^2$ and $r_m \geqslant \sum_{s=1}^{m}\alpha(s)^2$. Then for all $c \geqslant 0$, we have

$$\sup_{m\geqslant 1}(6t_m)^{-1}\exp\left[\frac{c(v_m + 2r_m)}{4t_m}\right]\mathbb{E}\left[S_m - \mathbb{E}[S_m] - \frac{3}{2}c(v_m + 2r_m)\right]_+ \leqslant 1$$

with $(a)_+ := (a \vee 0)$. □

We will use them to obtain concentration of sums of the shape $\sum_{s=m_1}^{m_2}(\phi_n(s)\lambda(s)^{-1}-\theta^\circ(s))^2$ and $\sum_{s=m_1}^{m_2}\phi_n(s)^2$.

We start by stating the set of assumptions which allow us to obtain our results.

Assumption 15 Suppose that λ is monotonically and polynomially decreasing, that is, there exist c in $[1,\infty[$ and a in $\mathbb{R}_+$ such that $\Lambda(m) \approx m^{-2a}$.

This assumption assures that $\Lambda_+(m) = \Lambda(m)$ for any m and that there exist a constant $L := L(\lambda)$ in $[1,\infty[$, independent of θ° such that for any sequence $(m_n)_{n\in\mathbb{N}^\star}$

$$\sup_{n\in\mathbb{N}^\star} m_n\Lambda(m_n)(n\Phi_n^{m_n})^{-1} \leqslant \sup_{n\in\mathbb{N}^\star}\Lambda(m_n)/\Lambda_\circ(m_n) \leqslant L.$$

It basically requires that we are in the situation **[o-o]** or **[s-o]** and is not valid under **[o-s]**.

Assumption 16 Let θ° and λ be such that there exists n° in $\mathbb{N}^\star$

$$0 < \kappa^\circ := \kappa^\circ(\theta^\circ,\lambda) := \inf_{n\geqslant n^\circ}\left\{(\Phi_n^\circ(\theta^\circ))^{-1}\left[\mathfrak{b}_{m_n^\circ}\wedge n^{-1}m_n^\circ\Lambda_\circ(m_n^\circ)\right]\right\} \leqslant 1$$

Assumption 17 Let $\mathfrak{a}$ and λ be sequences such that there exists $n^\star$ in $\mathbb{N}^\star$

$$0 < \kappa^\star := \kappa^\star(\mathfrak{a},\lambda) := \inf_{n>n^\star}\left\{(\Phi_n^\star)^{-1}\left[\mathfrak{a}_{m_n^\star}\wedge n^{-1}m_n^\star\Lambda_\circ(m_n^\star)\right]\right\} \leqslant 1.$$

The corollaries hereafter generalise the results obtained in Johannes et al. (2014) to finitely iterated posterior distributions. The proofs are sensibly similar to the original ones and we hence skip them.

Corollary 2.4.3.
Under Assumption 15 and Assumption 16, if, in addition $\log(G_n)/m_n^\circ \to 0$ as $n\to\infty$ then with $D^\circ := D^\circ(\theta^\circ,\lambda) = \lceil 5L/\kappa^\circ\rceil$ and $K^\circ := 10(2\vee\|\theta^\circ\|_{l^2}^2)L^2(16\vee D^\circ\Lambda_{D^\circ})$ we have, for any η $(1\leqslant\eta<\infty)$:

$$\lim_{n\to\infty}\mathbb{E}\left[\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}\left((K^\circ)^{-1}\Phi_n^\circ(\theta^\circ)\leqslant\|\theta^\circ-\boldsymbol{\theta}_{\overline{M}}\|_{l^2}^2\leqslant K^\circ\Phi_n^\circ(\theta^\circ)\right)\right]=1.$$

□

Corollary 2.4.4.
Under Assumption 15 and Assumption 17, if, in addition, $\log(G_n)/m_n^\star\to 0$ as $n\to\infty$ then, for any η $(1\leqslant\eta<\infty)$

- *for all θ° in $\Theta_{\mathfrak{a}}(r)$, with $D^\star := D^\star(\mathfrak{a},\lambda) = \lceil 5L/\kappa^\star\rceil$ and $K^\star := 16L^2(2\vee r)(16\vee D^\star\Lambda_{D^\star})$, we have*

$$\lim_{n\to\infty}\mathbb{E}\left[\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}\left(\|\theta^\circ-\boldsymbol{\theta}_{\overline{M}}\|_{l^2}^2\leqslant K^\star\Phi_n^\star\right)\right]=1;$$

- *for any monotonically increasing and unbounded sequence K_n holds*

$$\lim_{n\to\infty}\inf_{\theta^\circ\in\Theta_{\mathfrak{a}}(r)}\mathbb{E}\left[\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}\left(\|\theta^\circ-\boldsymbol{\theta}_{\overline{M}}\|_{l^2}^2\leqslant K_n\Phi_n^\star\right)\right]=1.$$

□

However, the following theorem assert that the results hold true in the asymptotic case where η tends to ∞. The proofs are displayed in appendix A.1 and appendix A.2 respectively.

Theorem 2.4.2.
Under Assumption 15, Assumption 16 and the condition that $\limsup_{n\to\infty} \log(G_n)(m_n^\circ)^{-1}$, *define* $D^\circ := \lceil 3(\kappa^\circ)^{-1} + 1 \rceil$ *and* $K^\circ := 16L \cdot [9 \vee D^\circ \Lambda_{D^\circ}]$*; then, we have for all* θ° *in* Θ,

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}\left((K^\circ)^{-1}\,\Phi_n^\circ(\theta^\circ) \leqslant \|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 \leqslant K^\circ \Phi_n^\circ(\theta^\circ)\right)\right] = 1.$$

Theorem 2.4.3.
Under Assumption 15, Assumption 17 and the condition that $\limsup_{n\to\infty} \frac{\log(G_n)}{m_n^\star}$, *define* $D^\star := \left\lceil \frac{3(1\vee r)}{\kappa^\star L} + 1 \right\rceil$ *and* $K^\star := 6(1 \vee r)(9L \vee D^\star \Lambda_{D^\star})$*; then, we have for all* θ° *in* $\Theta^{\mathfrak{a}}(r)$,

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 \leqslant K^\star \Phi_n^\star\right)\right] = 1,$$

and, for any increasing sequence K_n *such that* $\lim_{n\to\infty} K_n = \infty$,

$$\lim_{n\to\infty} \inf_{\theta^\circ \in \Theta^{\mathfrak{a}}(r)} \mathbb{E}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}\left(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 \leqslant K_n \Phi_n^\star\right)\right] = 1.$$

We have hence showed that the self informative Bayes carrier contracts around the true parameter with the oracle optimal rate of sieve priors and with minimax optimal rate over Sobolev's ellipsoids. We will see in section 3.3 that the self informative limit also converges with optimal rates.

2.5 On the shape of the posterior mean

We have hence seen that in a general case, considering the asymptotic iteration , the posterior distribution using a sieve prior contracts around the projection estimator and while using a hierarchical prior, the posterior contracts around some penalised contrast maximiser projection estimator.

It is also interesting to note that for any number of iteration η, the posterior mean can be written both as a shrinkage and as an aggregation estimator. Indeed, we have

$$\begin{aligned}
\mathbb{E}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left[\boldsymbol{\theta}_{\overline{M}}\right] &= \mathbb{E}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left[\sum\nolimits_{|m|\leqslant G} \boldsymbol{\theta}_{\overline{M}} \mathbb{1}_{M=m}\right] \\
&= \sum\nolimits_{|m|\leqslant G} \mathbb{E}^{\eta}_{\boldsymbol{\theta}_M|Y^n}\left[\boldsymbol{\theta}_M \mathbb{1}_{M=m}\right] \\
&= \sum\nolimits_{|m|\leqslant G} \mathbb{P}^{(\eta)}_{M|Y^n}(m = M)\, \mathbb{E}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left[\boldsymbol{\theta}_{\overline{m}}\right];
\end{aligned}$$

and we see here the aggregation form of this estimator.

On the other hand, if we write the expectation of the components individually, we obtain:

$$\begin{aligned}\mathbb{E}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left[\boldsymbol{\theta}_{\overline{M}}(s)\right] &= \mathbb{E}^{(\eta)}_{\boldsymbol{\theta}_{\overline{M}}|Y^n}\left[\boldsymbol{\theta}_{\overline{m}}(s)\mathbb{1}_{M\geqslant|s|}\right]\\ &= \mathbb{P}^{(\eta)}_{M|Y^n}(M\geqslant|s|)\,\mathbb{E}^{(\eta)}_{\boldsymbol{\theta}_{\overline{m}}|Y^n}\left[\boldsymbol{\theta}_{\overline{m}}(s)\right];\end{aligned}$$

where we see the shrinkage property.
Aggregation estimates gathered a lot of interest, see for example Rigollet and Tsybakov (2007). While considering such estimators, the goal is to reach the convergence rate of the best estimator contributing to the aggregation. In the next chapter, we hence investigate the properties of this estimator both in inverse Gaussian sequence space model and circular density deconvolution.

Chapter 3

Minimax and oracle optimal adaptive aggregation

We inquire in this chapter the properties of aggregation estimators as introduced in section 2.5. We introduce first a skim of proof for oracle and minimax optimality of this kind of estimator before applying it to the inverse Gaussian sequence space and the circular deconvolution models respectively introduced in section 1.5 and section 1.6, including in presence of dependance and partially known operator.

Remind that we are interested in the estimation of an element f of a Hilbert space $(\Xi, \langle\cdot|\cdot\rangle_{\Xi})$, in an optimal manner with respect to the norm $\|\cdot\|_{\Xi}$ induced by the inner product. Considering an index set $\mathbb{F} = \mathbb{Z} \,\text{or}\, \mathbb{N}$; an orthonormal system $(e_s)_{s\in\mathbb{F}}$ in $(\Xi, \langle\cdot|\cdot\rangle_{\Xi})$; and the space of $\mathbb{F}$-indexed, $\mathbb{K}(= \mathbb{C} \,\text{or}\, \mathbb{R})$-valued sequences Θ we defined the generalised Fourier transform $\mathcal{F} : \Xi \to \Theta, x \mapsto [x] = (\langle x|e_s\rangle_{\Xi})_{s\in\mathbb{F}}$.

We then let T be a linear operator from Ξ to itself such that, for any s in $\mathbb{F}$, there exists $\lambda(s)$ in $\mathbb{K} \setminus \{0\}$ such that $T(e_s) = \lambda(s)e_s$ and we denote $g := T(f)$; $h := \mathcal{F}^{\star}((\lambda(s))_{s\in\mathbb{F}})$; $\theta^{\circ} := \mathcal{F}(f)$; and $\phi := \mathcal{F}(g)$.

Under Assumption 3, we define a strictly stationary stochastic process $Y = (Y_p)_{p\in\mathbb{Z}}$ in which for any p in $\mathbb{Z}$, Y_p is a Ξ indexed stochastic process verifying, for any x and y in Ξ, $\mathbb{E}[Y_p(x)] = \langle g|x\rangle_{\Xi}$ and $\mathbb{C}\mathrm{ov}(Y_p(x), Y_p(y)) = \langle x|y\rangle_{\Xi}$. In particular for any s and s' in $\mathbb{F}$ we have $\mathbb{E}[Y_p(e_s)] = \phi(s)$ and $\mathbb{C}\mathrm{ov}(Y_p(e_s), Y_p(e_{s'})) = \mathbb{1}_{\{s=s'\}}$. We then assume to observe the sub-vector $Y^n = (Y_p)_{p\in[\![1,n]\!]}$ of Y. Under Assumption 4, in addition to observing Y^n, we define a strictly stationary stochastic process $\varepsilon = (\varepsilon_p)_{p\in\mathbb{Z}}$ in which for any p in $\mathbb{Z}$, ε_p is a Ξ indexed stochastic process verifying, for any x and y in Ξ, $\mathbb{E}[\varepsilon_p(x)] = \langle h|x\rangle_{\Xi}$ and $\mathbb{C}\mathrm{ov}(\varepsilon_p(x), \varepsilon_p(y)) = \langle x|y\rangle_{\Xi}$. In particular for any s and s' in $\mathbb{F}$ we have $\mathbb{E}[\varepsilon_p(e_s)] = \lambda(s)$ and $\mathbb{C}\mathrm{ov}(\varepsilon_p(e_s), \varepsilon_p(e_{s'})) = \mathbb{1}_{\{s=s'\}}$. We then assumed to observe the sub-vector $\varepsilon^{n_\lambda} = (\varepsilon_p)_{p\in[\![1,n_\lambda]\!]}$ of ε.

Then, we pointed out that, for each s, a naive estimator for $\phi(s)$ is $\phi_n(s) = n^{-1}\sum_{p=1}^{n} Y_p(e_s)$, and hence an estimator for $\theta^{\circ}(s)$ could be, under Assumption 3, $\theta_n(s) = \phi_n(s)\lambda(s)^{-1}$ as we assumed $\lambda(s) \neq 0$, and, under Assumption 4, we define $\lambda_{n_\lambda}(s) = n_\lambda^{-1}\sum_{p=1}^{n_\lambda} \varepsilon_p(e_s)$, $\lambda_{n_\lambda}^{+}(s) = \lambda_{n_\lambda}(s)^{-1}\mathbb{1}_{\{\lambda_{n_\lambda}(s)>n_\lambda^{-1}\}}$, and $\theta_{n,n_\lambda}(s) = \phi_n(s)\lambda_{n_\lambda}^{+}(s)$. Defining the sieve family $(\mathbb{F}_m)_{m\in\mathbb{N}} = (\{s \in \mathbb{F} : |s| \leqslant m\})_{m\in\mathbb{N}}$, we then introduced the families of projection estimators $(\phi_{n,\overline{m}})_{m\in\mathbb{N}}) = ((\phi_n(s)\mathbb{1}_{\{|s|\leqslant m\}})_{s\in\mathbb{F}})_{m\in\mathbb{N}}$, and $(\theta_{n,\overline{m}})_{m\in\mathbb{N}} = ((\theta_n(s)\mathbb{1}_{\{|s|\leqslant m\}})_{s\in\mathbb{F}})_{m\in\mathbb{N}}$ under Assumption 3 or $(\theta_{n,n_\lambda,\overline{m}})_{m\in\mathbb{N}} = ((\theta_{n,n_\lambda}(s)\mathbb{1}_{\{|s|\leqslant m\}})_{s\in\mathbb{F}})_{m\in\mathbb{N}}$ under Assumption 4.

We then presented the oracle and minimax risk for those projection estimators and highlighted that taking their inverse Fourier transform would give an estimator of f with the

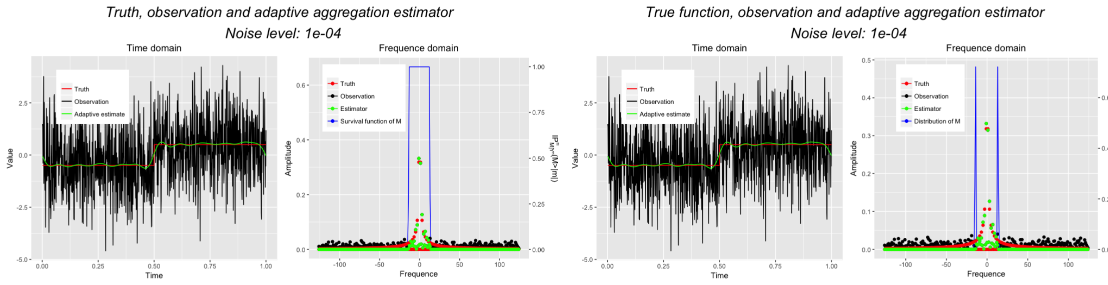

Figure 3.1: Aggregation estimator on an Gaussian sequence space model, direct problem case

same quadratic risk, thanks to Plancherel theorem. We also highlighted that the optimal choice for the threshold parameter m depends either on θ° itself or on its regularity class in the minimax case. This dependence justifies the need for data-driven methods such as the penalised contrast minimisation. Interestingly, we saw in the last chapter that this method could be seen, from the Bayesian point of view, as the self informative limit of a family of hierarchical sieve priors and used this fact to prove its optimality in terms of speed of convergence in probability. The posterior mean of those hierarchical sieve priors has been expressed both as an aggregation and as a shrinkage estimator and we will mimic this structure in this chapter. This allows us to suggest a proof strategy for optimality of such estimators and apply it to our two example models.

3.1 Shape of the aggregation estimators

We want to define a family of estimators $(\widehat{\theta}^{(\eta)})_{\eta\in]0,\infty]}$ of θ° and estimators for f, $\widehat{f}^{(\eta)} := \mathcal{F}^{\star}(\widehat{\theta}^{(\eta)})$ such that, for any η, $\widehat{\theta}^{(\eta)}$ has the shape

$$(\widehat{\theta}^{(\eta)}(s))_{s\in\mathbb{F}} = \sum\nolimits_{m\in\mathbb{N}} \mathbb{P}_{M}^{(\eta)}(m)\cdot(\theta_{n,\overline{m}}(s))_{s\in\mathbb{F}} = \sum\nolimits_{m\geqslant|s|} \mathbb{P}_{M}^{(\eta)}(m)\cdot(\theta_{n}(s))_{s\in\mathbb{F}} \tag{3.1}$$

under Assumption 3, and

$$(\widehat{\theta}^{(\eta)}(s))_{s\in\mathbb{F}} = \sum\nolimits_{m\in\mathbb{N}} \widehat{\mathbb{P}}_{M}^{(\eta)}(m)\cdot(\theta_{n,n_{\lambda},\overline{m}}(s))_{s\in\mathbb{F}} = \sum\nolimits_{m\geqslant|s|} \widehat{\mathbb{P}}_{M}^{(\eta)}(m)\cdot(\theta_{n,n_{\lambda}}(s))_{s\in\mathbb{F}} \tag{3.2}$$

under Assumption 4. The sequence $(\mathbb{P}_{M}^{(\eta)}(m))_{m\in\mathbb{N}}$ is the aggregation sequence. Under Assumption 3 it depends on the observations Y^{n} as well on the known operator T through its eigen values $(\lambda(s))_{s\in\mathbb{F}}$ whereas under Assumption 4, it only depends on the observed data Y^{n} and $\varepsilon^{n_{\lambda}}$. This notation is motivated by the Bayesian inspiration of the method.

We give in fig. 3.1 an illustration of the aggregation estimator used in a Gaussian sequence space model in the direct problem case, that is to say $\lambda(s) = 1$ for any s in $\mathbb{N}$.

Taking inspiration in the posterior distributions obtained with a hierarchical prior in the previous chapter, we will give the following shape to the aggregation weights. In the case

Assumption 3 let be the following functions

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^{\Lambda} : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \mathrm{pen}^{\Lambda}(m);$$
$$\mathbb{P}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \frac{\exp[-\eta n(-\Upsilon(m)+\mathrm{pen}^{\Lambda}(m))]}{\sum_{k=0}^{n} \exp[-\eta n(-\Upsilon(k)+\mathrm{pen}^{\Lambda}(k))]} \mathbb{1}_{m \leqslant n}; \quad (3.3)$$

where Υ depends on the observations Y^n as well as the known operator T through the sequence λ of its eigen values; and pen^{Λ} depends only on the parameter T through the sequence λ of its eigen values. Under Assumption 4, we define

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^{\widehat{\Lambda}} : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \mathrm{pen}^{\widehat{\Lambda}}(m);$$
$$\widehat{\mathbb{P}}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+; \quad m \mapsto \frac{\exp[\eta n(\Upsilon(m)-\mathrm{pen}^{\widehat{\Lambda}}(m))]}{\sum_{k=0}^{n} \exp[\eta n(\Upsilon(k)-\mathrm{pen}^{\widehat{\Lambda}}(k))]} \mathbb{1}_{m \leqslant n}; \quad (3.4)$$

where Υ depends solely the observations Y^n and ε^{n_λ}; and $\mathrm{pen}^{\widehat{\Lambda}}$ depends only on the observations ε^{n_λ}. The functions Υ, and pen^{Λ} will respectively be called contrast and penalty. For any subset S of $\mathbb{N}$, we denote $\mathbb{P}_M^{(\eta)}(S) = \sum_{k \in S} \mathbb{P}_M^{(\eta)}(k)$. One would expect that as the amount of data increases, the number of coefficients estimated increases too, as our observations allows us to recover more information about the system of interest as illustrated in fig. 3.2 by representing $\mathbb{P}_M^{(\eta)}(\llbracket m, n \rrbracket)$ for increasnig values of n.

Consider first the asymptotic when one lets η tend to infinity. Under Assumption 3, following a model selection approach (c.f. Barron et al. (1999) and Massart (2007) for an extensive description), a dimension parameter $\widehat{m}$ is determined among a collection of admissible values $\llbracket 1, n \rrbracket$ by minimising the penalised contrast function $-\|\theta_{n,\overline{m}}\|_{l^2} + \mathrm{pen}^{\Lambda}(m)$, that is

$$\widetilde{m} := \arg\min_{m \in \llbracket 1, n \rrbracket} \big\{ -\Upsilon(m) + \mathrm{pen}^{\Lambda}(m) \big\}. \quad (3.5)$$

If $\widetilde{m}$ minimises uniquely the penalised contrast function, then it is easily seen that the discrete probability measure $\mathbb{P}_M^{(\eta)}$ on the set $\llbracket 1, n \rrbracket$ given by the weights $\mathbb{P}_M^{(\eta)}(\{m\}) = \mathbb{P}_M^{(\eta)}(m)$ as in (3.3) degenerates to a Dirac measure $\delta_{\widetilde{m}}$ on the point $\widetilde{m}$ as $\eta \to \infty$. Precisely, for any $m \in \llbracket 1, n \rrbracket$ holds

$$\lim_{\eta \to \infty} \mathbb{P}_M^{(\eta)}(m) = \delta_{\widetilde{m}}(\{m\}) =: \mathbb{P}_M^{(\infty)}(m) \quad (3.6)$$

Thereby, in the sequel we consider the model selected estimator

$$\theta_{n,\overline{\widetilde{m}}} = \widehat{\theta}^{(\infty)} = \sum_{m \in \llbracket 1, n \rrbracket} \mathbb{P}_M^{(\infty)}(m) \theta_{n,\overline{m}}$$

as an aggregation with respect to the discrete measure $\mathbb{P}_M^{(\infty)} = \delta_{\widetilde{m}}$ on the set $\llbracket 1, n \rrbracket$. We give in fig. 3.3 an illustration of the model selection estimator used on a Gaussian sequence space model, in the direct problem case, that is to say $\lambda(s) = 1$ for all s.
Under Assumption 4 consider again a model selection approach by minimising now the penalised contrast function $\Upsilon(m) + \mathrm{pen}^{\widehat{\Lambda}}(m)$, that is

$$\widehat{m} := \arg\min_{m \in \llbracket 1, n \rrbracket} \big\{ -\Upsilon(m) + \mathrm{pen}^{\widehat{\Lambda}}(m) \big\}. \quad (3.7)$$

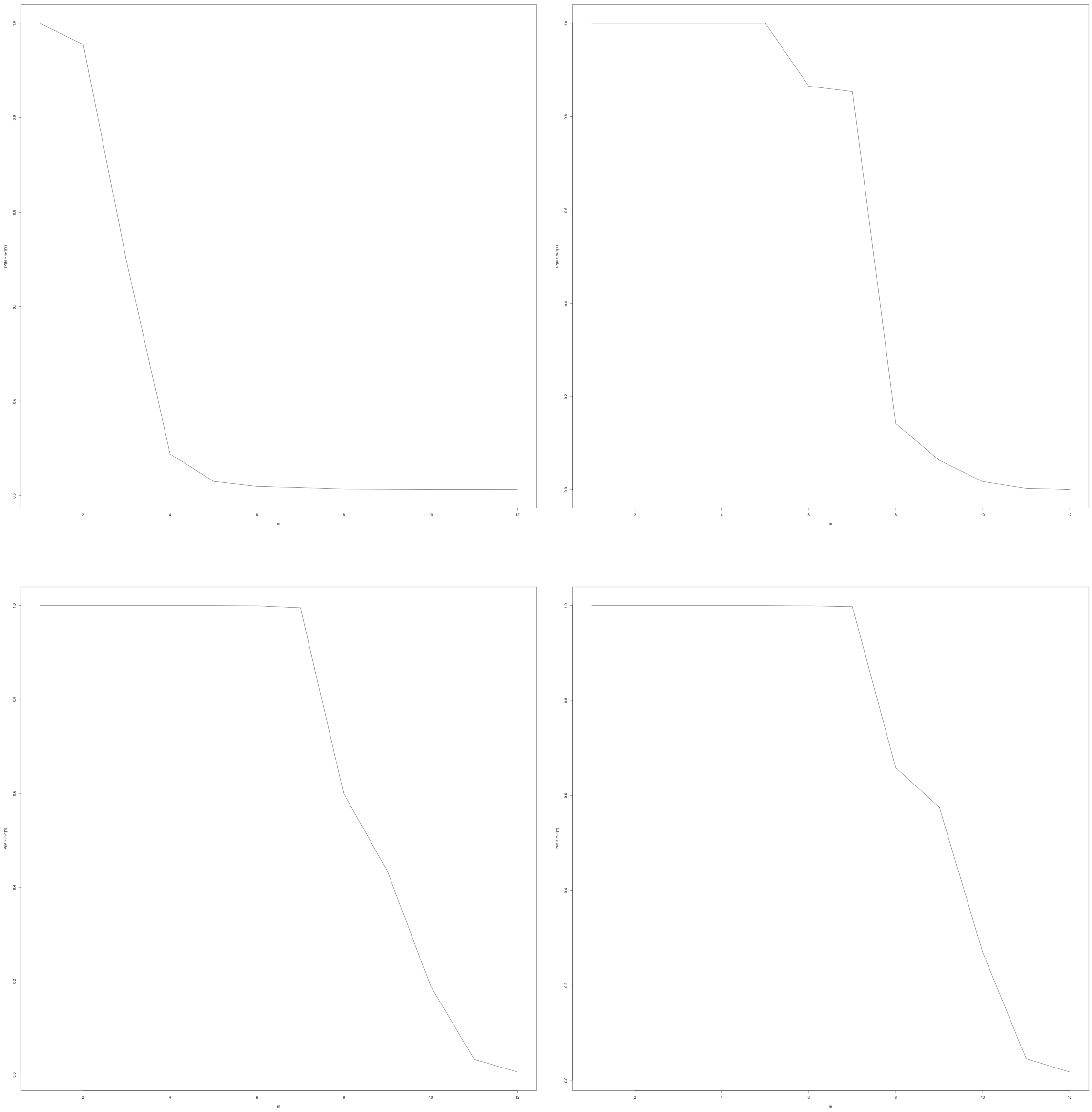

Figure 3.2: Evolution of the aggregation weights

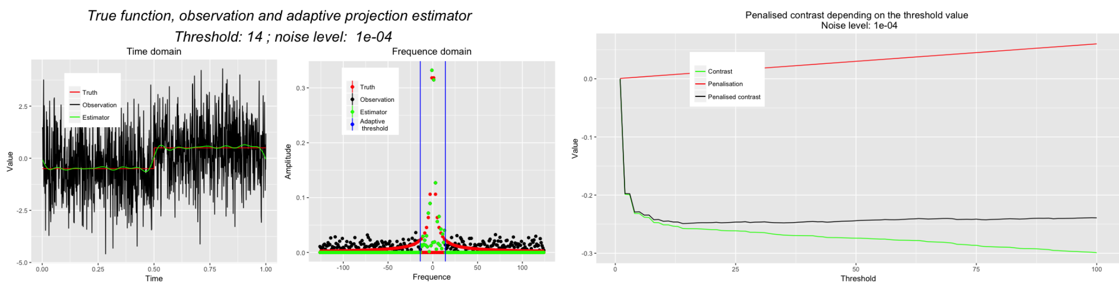

Figure 3.3: Model selection estimator on an Gaussian sequence space model, direct problem case

If $\widehat{m}$ minimises uniquely the penalised contrast function, then for any $m \in [\![1,n]\!]$ holds

$$\lim_{\eta\to\infty} \widehat{\mathbb{P}}_M^{(\eta)}(m) = \delta_{\widehat{m}}(\{m\}) =: \widehat{\mathbb{P}}_M^{(\infty)}. \tag{3.8}$$

Thereby, we consider again the model selected estimator $\theta_{n,\overline{\widehat{m}}} = \widehat{\theta}^{(\infty)} = \sum_{m\in[\![1,n]\!]} \widehat{\mathbb{P}}_M^{(\infty)}(m)\theta_{n,n_\lambda,\overline{m}}$ as an aggregation with respect to the discrete measure $\mathbb{P}_M^{(\infty)} = \delta_{\widehat{m}}$ on the set $[\![1,n]\!]$.
We will consider two examples in this chapter, namely the inverse Gaussian sequence space model as well as the circular deconvolution model. In both cases the functions Υ, pen^{Λ}, and $\mathrm{pen}^{\widehat{\Lambda}}$ take the same shape which we hence give here.

Definition 37 Under Assumption 3, let be a universal constant κ to be fixed depending on the considered model. For any m in $[\![1,n]\!]$, remind that $\Lambda(m) = |\lambda(m)|^{-2}$, and $\Lambda_+(m) = \max\{\Lambda(s), s \in \mathbb{F}_m\}$ and define

$$\begin{aligned}
\Upsilon(m) &:= \|\theta_{n,\overline{m}}\|_{l^2}^2; & \delta_\Lambda(m) &:= \tfrac{\log^2(m\Lambda_+(m)\vee(m+2))}{\log^2(m+2)} \geqslant 1;\\
\Delta_\Lambda(m) &:= \delta_\Lambda(m) m\Lambda_+(m); & \mathrm{pen}^\Lambda(m) &:= \kappa\, \Delta_\Lambda(m)\, n^{-1}.
\end{aligned}$$

□

Definition 38 Under Assumption 4, let be a universal constant κ to be fixed depending on the considered model. Then, for any m in $\mathbb{N}$, we define

$$\begin{aligned}
\Upsilon(m) &:= \|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}^2; & \widehat{\Lambda}(s) &:= |\lambda_{n_\lambda}^+(s)|^2\\
\widehat{\Lambda}_+(m) &:= \max\{\widehat{\Lambda}(l), l \in [\![1,m]\!]\}; & \delta_{\widehat{\Lambda}}(m) &:= \tfrac{\log^2(m\widehat{\Lambda}_+(m)\vee(m+2))}{\log^2(m+2)} \geqslant 1;\\
\Delta_{\widehat{\Lambda}}(m) &:= \delta_{\widehat{\Lambda}}(m) m\widehat{\Lambda}_+(m); & \mathrm{pen}^{\widehat{\Lambda}}(m) &:= \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}.
\end{aligned}$$

□

Notice that, with the exception of the constant κ, our estimator is now fully determined, in both cases Assumption 3 and Assumption 4.

3.2 Strategy of proof for optimality of aggregation estimator

As we have now given a precise shape to our aggregation estimator, we propose a strategy to compute upper bounds for its convergence rate in l^2-norm. Our method is inspired by the strategy to compute upper bounds for the contraction rate of hierarchical sieves we presented in the previous chapter. We will hence highlight a decomposition of the risk which separates the risk obtained by taking values of the threshold which are respectively "too small", "too large", or "optimal". Those terms should be understood with respect to the quadratic risk of the projection estimator associated with this choice of threshold. One would then prove that the values of the threshold which are too small or too large do not receive an important weight under $\mathbb{P}_M^{(\eta)}$ or $\widehat{P}_M^{(\eta)}$. Before going any further, notice that, for any m and $m^\bullet$ in $\mathbb{N}$, the aggregation weights can be bounded in the following way:

$$\frac{\exp[-\eta n(-\|\theta_{n,\overline{m}}\|_{l^2}+\mathrm{pen}^\Lambda(m))]}{\sum_{k=0}^{n}\exp[-\eta n(-\|\theta_{n,\overline{k}}\|_{l^2}+\mathrm{pen}^\Lambda(k))]}\mathbb{1}_{m\leqslant n}$$
$$\leqslant \exp[-\eta n(\|\theta_{n,\overline{m}}\|_{l^2}-\|\theta_{n,\overline{m^\bullet}}\|_{l^2}+\mathrm{pen}^\Lambda(m^\bullet)-\mathrm{pen}^\Lambda(m))]\mathbb{1}_{m\leqslant n};\text{ and}$$

$$\frac{\exp[-\eta n(-\|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}+\mathrm{pen}^\Lambda(m))]}{\sum_{k=0}^{n}\exp[-\eta n(-\|\theta_{n,n_\lambda,\overline{k}}\|_{l^2}+\mathrm{pen}^\Lambda(k))]}\mathbb{1}_{m\leqslant n}$$
$$\leqslant \exp[-\eta n(\|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}-\|\theta_{n,n_\lambda,\overline{m^\bullet}}\|_{l^2}+\mathrm{pen}^\Lambda(m^\bullet)-\mathrm{pen}^\Lambda(m))]\mathbb{1}_{m\leqslant n}.$$

Then, the following lemma, which proof is given in appendix B allows to derive an upper bound which is easier to control.

Lemma 3.2.1.
Given $n\in\mathbb{N}$ and $\theta^\circ,\breve{\theta}\in l_2$ consider the families of orthogonal projections $\{\breve{\theta}_{\overline{m}}=\Pi_{\overline{m}}\breve{\theta}, m\in[\![1,n]\!]\}$ and $\{\theta^\circ_{\overline{m}}=\Pi_{\overline{m}}\theta^\circ, m\in[\![1,n]\!]\}$.
If $\|\Pi_{\overline{m}}^\perp\theta^\circ\|_{l^2}^2=\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_m^2(\theta^\circ)$ for all $m\in[\![1,n]\!]$, then for any $l\in[\![1,n]\!]$ holds

(i) $\|\breve{\theta}_{\overline{k}}\|_{l^2}^2-\|\breve{\theta}_{\overline{l}}\|_{l^2}^2\leqslant\frac{11}{2}\|\breve{\theta}_{\overline{l}}-\theta^\circ_{\overline{l}}\|_{l^2}^2-\frac{1}{2}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\{\mathfrak{b}_k^2(\theta^\circ)-\mathfrak{b}_l^2(\theta^\circ)\}$, *for all $k\in[\![1,l[\![$;*

(ii) $\|\breve{\theta}_{\overline{k}}\|_{l^2}^2-\|\breve{\theta}_{\overline{l}}\|_{l^2}^2\leqslant\frac{7}{2}\|\breve{\theta}_{\overline{k}}-\theta^\circ_{\overline{k}}\|_{l^2}^2+\frac{3}{2}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\{\mathfrak{b}_l^2(\theta^\circ)-\mathfrak{b}_k^2(\theta^\circ)\}$, *for all $k\in]\!]l,n]\!]$.*

□

3.2.1 Known operator

Consider first the case Assumption 3. We shall hence keep in mind 3.1, 3.3, Definition 37 as well as 3.5 and 3.6. **Note that the detailed proofs for all results given here can be found in appendix B.1.**
Both for the quadratic and the maximal risk, our strategy is based on the decomposition of the quadratic loss function displayed in Lemma 3.2.2. This decomposition is independent of the model and only relies on the fact that the parameter space is equipped with a nested sieve and the fact that our estimator aggregation structure takes advantage of it.

Lemma 3.2.2.
First writing the l^2-distance between θ° and $\widehat{\theta}^\eta$ we obtain, for any m_- and m_+ in $[\![1,n]\!]$ such that $m_-\leqslant m_+$, and sequence $(\mathrm{pen}(m))_{m\in\mathbb{N}}$ of compensating terms,

$$\begin{aligned}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2 &\leqslant \tfrac{2}{7}\,\mathrm{pen}(m_+)+2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ &+2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathbb{P}_M^{(\eta)}([\![1,m_-[\![)+\tfrac{2}{7}\sum_{m\in]\!]m_+,n]\!]}\mathrm{pen}(m)\,\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2<\mathrm{pen}(m)/7\}}\\ +2\sum_{m\in[\![m_+,n]\!]}\left(\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2-\mathrm{pen}(m)/7\right)_+ &+\tfrac{2}{7}\sum_{m\in]\!]m_+,n]\!]}\mathrm{pen}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2\geqslant\mathrm{pen}(m)/7\}}.\end{aligned}\tag{3.9}$$

□

The proof strategy will be articulated around the search for sequences m_+, m_- and $\text{pen}(m)$ such that each term is properly controlled. In practice, the terms $\frac{2}{7}\,\text{pen}(m_+)$ and $2\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m_-}(\theta^\circ)$ will be the leading terms in the sum.

3.2.1.1 Quadratic risk bounds

We propose a strategy which allows to prove that the sequence defined hereafter is an upper bound for the quadratic risk of the aggregation estimator we just defined.

Definition 39 Remind that we defined for any θ in Θ and m in $\mathbb{N}$ the following term $\mathfrak{b}^2_m(\theta) = \|\theta_{\underline{m}}\|^2_{l^2}\|\theta_{\underline{0}}\|^{-2}_{l^2} \leqslant 1$. We then define a family of sequences $(\mathfrak{R}^m_n(\theta^\circ))_{m\in\mathbb{N}} := (\mathfrak{R}^m_n(\theta^\circ,\Lambda))_{m\in\mathbb{N}} = ([\mathfrak{b}^2_m(\theta^\circ) \vee \text{pen}^\Lambda(m)/\kappa])_{m\in\mathbb{N}}$ and hence it holds for all m in $[\![1,n]\!]$

$$[\|\theta^\circ_{\underline{0}}\|^2_{l^2}+\kappa]\mathfrak{R}^m_n(\theta^\circ) \geqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ) \vee \text{pen}^\Lambda(m). \tag{3.10}$$

We intend to prove that the specific choice

$$m^\dagger_n(\theta^\circ) := \arg\min\{\mathfrak{R}^m_n(\theta^\circ), m\in\mathbb{N}\} \in [\![1,n]\!];$$
$$\mathfrak{R}^\dagger_n(\theta^\circ) := \mathfrak{R}^\dagger_n(\theta^\circ,\Lambda) := \min\{\mathfrak{R}^m_n(\theta^\circ), m\in\mathbb{N}\}$$

with $\mathfrak{R}^{m^\dagger_n}_n(\theta^\circ,\Lambda) = \mathfrak{R}^\dagger_n(\theta^\circ,\Lambda)$ defines an upper bound for the convergence rate of the aggregation estimators. □

Note that the proofs for the results displayed here can be found in appendix B.1.1

Remark 3.2.1 *The following statements can be shown using the same arguments as in Remark 1.3.1 by exploiting that the sequence $\mathfrak{b}^2_m(\theta^\circ)$ is non-increasing with limit zero and $\mathfrak{b}^2_0(\theta^\circ) \leqslant 1$. By construction for all $n\in\mathbb{N}$ it hold $\mathfrak{R}^\dagger_n(\theta^\circ) \geqslant n^{-1}$ and $\mathfrak{R}^\dagger_n(\theta^\circ) = \mathfrak{o}_n(1)$. Moreover, for all $n\in\mathbb{N}$ we have $m^\dagger_n(\theta^\circ)\in[\![1,n]\!]$, $m^\dagger_n(\theta^\circ) = \arg\min\{\mathfrak{R}^m_n(\theta^\circ), m\in[\![1,n]\!]\}$ and $\mathfrak{R}^\dagger_n(\theta^\circ) = \min\{\mathfrak{R}^m_n(\theta^\circ), m\in[\![1,n]\!]\}$. Thereby, in case (p) we conclude that $m^\dagger_n(\theta^\circ) = K$ and the rate $\mathfrak{R}^\dagger_n(\theta^\circ)$ is parametric, that is $\mathfrak{R}^\dagger_n(\theta^\circ) = \Delta_\Lambda(K)n^{-1} \approx n^{-1}$, and hence equals the oracle rate $\mathcal{R}^\circ_n(\theta^\circ)$, i.e. $\mathcal{R}^\circ_n(\theta^\circ) \approx \mathfrak{R}^\dagger_n(\theta^\circ)$. On the other hand side, in case (np) the rate $\mathfrak{R}^\dagger_n(\theta^\circ)$ is nonparametric, that is, $n\mathfrak{R}^\dagger_n(\theta^\circ)\to\infty$ and $m^\dagger_n(\theta^\circ)\to\infty$ as $n\to\infty$. Moreover, by construction holds $\mathfrak{R}^\dagger_n(\theta^\circ) \geqslant \mathcal{R}^\circ_n(\theta^\circ)$.* □

Let us first briefly illustrate the last definitions by stating the order of $m^\dagger_n(\theta^\circ)$ and $\mathfrak{R}^\dagger_n(\theta^\circ)$ in the cases considered in Num. discussion 1.3.1

Numerical discussion 3.2.1.
Let us illustrate Definition 39 considering as in Num. discussion 1.3.1 usual behaviour **[o-o]**, **[s-o]** and **[o-s]** for the sequences $(\mathfrak{b}_m(\theta^\circ))_{m\in\mathbb{N}}$ and $(\Lambda(m))_{m\in\mathbb{N}}$:

[o-o] Since $\mathfrak{b}^2_m(\theta^\circ) \approx m^{-2p}$ and $\Delta_\Lambda(m) \approx m^{2a+1}$ follows $\mathfrak{R}^{m^\dagger_n}_n(\theta^\circ,\Lambda) \approx (m^\dagger_n)^{-2p} \approx \Delta_\Lambda(m^\dagger_n)n^{-1} \approx (m^\dagger_n)^{2a+1}n^{-1}$ which implies $m^\dagger_n \approx n^{1/(2p+2a+1)}$, $\delta_\Lambda(m^\dagger_n)m^\dagger_n \approx n^{1/(2p+2a+1)}$, $\mathfrak{R}^\dagger_n(\theta^\circ) \approx n^{-2p/(2p+2a+1)}$ and $|\log\mathfrak{R}^\dagger_n(\theta^\circ)| \approx (\log n)$.

[o-s] Since $\mathfrak{b}^2_m(\theta^\circ) \approx m^{-2p}$ and $\Delta_\Lambda(m) \approx m^{1+4a}\exp(m^{2a})$ follows $\mathfrak{R}^{m^\dagger_n}_n(\theta^\circ,\Lambda) \approx (m^\dagger_n)^{-2p} \approx \Delta_\Lambda(m^\dagger_n)n^{-1} \approx (m^\dagger_n)^{1+4a}\exp((m^\dagger_n)^{2a})$ which implies $m^\dagger_n \approx (\log n)^{1/(2a)}$, $\delta_\Lambda(m^\dagger_n)m^\dagger_n \approx (\log n)^{2+1/(2a)}$, $\mathfrak{R}^\dagger_n(\theta^\circ) \approx (\log n)^{-p/a}$ and $|\log\mathfrak{R}^\dagger_n(\theta^\circ)| \approx (\log\log n)$.

[s-o] Since $\mathfrak{b}_m^2(\theta^\circ) \approx \exp(-m^{2p})$ and $\Delta_\Lambda(m) \approx m^{2a+1}$ follows $\mathfrak{R}_n^{m_n^\dagger}(\theta^\circ, \Lambda) \approx \exp(-(m_n^\dagger)^{2p}) \approx \Delta_\Lambda(m_n^\dagger)n^{-1} \approx (m_n^\dagger)^{2a+1}n^{-1}$ which implies $m_n^\dagger \approx (\log n)^{1/(2p)}$, $\delta_\Lambda(m_n^\dagger)m_n^\dagger \approx (\log n)^{1/(2p)}$, $\mathfrak{R}_n^\dagger(\theta^\circ) \approx (\log n)^{(2a+1)/(2p)}n^{-1}$ and $|\log \mathfrak{R}_n^\dagger(\theta^\circ)| \approx (\log n)$.

We note that in the three cases **[o-o]**, **[o-s]** and **[s-o]** the rate $\mathfrak{R}_n^\dagger(\theta^\circ)$ coincide with the oracle rate $\mathcal{R}_n^\circ(\theta^\circ)$ derived in Num. discussion 1.3.1 **[o-o]**, **[o-s]** and **[s-o]**, respectively. □

Under Definition 37 for arbitrary $m_+^\dagger, m_-^\dagger \in [\![1,n]\!]$ let us define

$$m_- := \min\left\{m \in [\![1, m_-^\dagger]\!] : \|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ) \leqslant [\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 4\kappa]\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right\} \quad \text{and}$$

$$m_+ := \max\left\{m \in [\![m_+^\dagger, n]\!] : \mathrm{pen}^\Lambda(m) \leqslant 2[3\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 2\kappa]\mathfrak{R}_n^{m_+^\dagger}(\theta^\circ)\right\} \quad (3.11)$$

where the defining set obviously contains $m_-^\dagger$ and $m_+^\dagger$, respectively, and hence, it is not empty.

Considering the third and fourth terms on the right hand side of (3.9), we will use the following lemma to control them.

Lemma 3.2.3.
Consider the data-driven aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3). Under Definition 37 with $\kappa \geqslant 8\log(3e)$ for any $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ and associated $m_+, m_- \in [\![1,n]\!]$ as in (3.11) hold

(i) $\mathbb{P}_M^{(\eta)}([\![1, m_-[\![)\mathbb{1}_{\left\{\|\theta_{\overline{n,m_-^\dagger}} - \theta^\circ_{\overline{m_-^\dagger}}\|_{l^2}^2 < \kappa\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)/7\right\}} \leqslant \frac{1}{\eta\kappa}\mathbb{1}_{\{m_->1\}}\exp\left(-\frac{3\eta\kappa}{14}n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right)$;

(ii) $\sum_{m\in]\!]m_+,n]\!]} \mathrm{pen}^\Lambda(m)\, \mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|_{l^2}^2 < \mathrm{pen}^\Lambda(m)/7\}} \leqslant n^{-1}\{\frac{16}{\kappa\eta^2} + \frac{8}{\eta}\}$.

□

We combine the upper bound in Lemma 3.2.2 and the bounds given in Lemma 3.2.3. Clearly, due to Lemma 3.2.3 we have

$$\mathbb{E}\,\mathbb{P}_M^{(\eta)}([\![1, m_-[\![) \leqslant \mathbb{1}_{\{m_->1\}}\{\tfrac{1}{\eta\kappa}\exp\left(-\tfrac{3\eta\kappa}{14}n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right) + \mathbb{P}\left(\|\theta_{\overline{n,m_-^\dagger}} - \theta^\circ_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \tfrac{\kappa}{7}\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right)\}$$

and, hence from (3.2.2) follows immediately

$$\begin{aligned}
\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 &\leqslant n^{-1}\{\tfrac{32}{7\kappa\eta^2} + \tfrac{16}{7\eta}\} + \tfrac{2}{\eta\kappa}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\left(-\tfrac{3\eta\kappa}{14}n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right) \\
&+ 2\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}\left(\|\theta_{\overline{n,m_-^\dagger}} - \theta^\circ_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \tfrac{\kappa}{7}\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right) + \tfrac{2}{7}\mathrm{pen}^\Lambda(m_+) + 2\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ) \\
&+ 2\sum_{m\in[\![m_+^\dagger,n]\!]} \mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|_{l^2}^2 - \tfrac{1}{7}\mathrm{pen}^\Lambda(m)\right)_+ \\
&+ \tfrac{2}{7}\sum_{m\in[\![m_+^\dagger,n]\!]} \mathrm{pen}^\Lambda(m)\,\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|_{l^2}^2 \geqslant \tfrac{1}{7}\mathrm{pen}^\Lambda(m)\right) \qquad (3.12)
\end{aligned}$$

The next result can be directly deduced from Lemma 3.2.3 by letting $\eta \to \infty$. However, we think the direct proof given in appendix B.1 provides an interesting illustration of the values $m_+, m_- \in [\![1,n]\!]$ as defined in (3.11).

Lemma 3.2.4.
Consider the data-driven model selection weights $\mathbb{P}_M^{(\infty)}$ *as in* (3.6). *Under definition Definition 37 for any* $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ *and associated* $m_+, m_- \in [\![1,n]\!]$ *as in* (3.11) *hold*

(i) $\mathbb{P}_M^{(\infty)}([\![1,m_-[\![)\mathbb{1}_{\{\|\theta_{\overline{n,m_-^\dagger}} - \theta^\circ_{\overline{m_-^\dagger}}\|^2_{l^2} < \kappa\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)/7\}} = 0;$

(ii) $\sum_{m\in]\!]m_+,n]\!]} \mathrm{pen}^\Lambda(m)\, \mathbb{P}_M^{(\infty)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} < \mathrm{pen}^\Lambda(m)/7\}} = 0.$

□

We combine again the upper bound in Lemma 3.2.2 and the bounds given in Lemma 3.2.4. Clearly, due to Lemma 3.2.4 we have $\mathbb{E}\,\mathbb{P}_M^{(\infty)}([\![1,m_-[\![) = \mathbb{P}\big(\|\theta_{\overline{n,m_-^\dagger}} - \theta^\circ_{\overline{m_-^\dagger}}\|^2_{l^2} \geqslant \kappa\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)/7\big)$ and, hence from (3.2.2) follows immediately

$$\mathbb{E}\|\theta_{n,\widehat{\overline{m}}} - \theta^\circ\|^2_{l^2} \leqslant 2\sum\nolimits_{m\in[\![m_+^\dagger,n]\!]} \mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} - \tfrac{1}{7}\,\mathrm{pen}^\Lambda(m)\right)_+$$
$$\tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+) + 2\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m_-}(\theta^\circ) + 2\|\theta^\circ_{\underline{0}}\|^2_{l^2}\mathbb{1}_{\{m_->1\}}\,\mathbb{P}\left(\|\theta_{\overline{n,m_-^\dagger}} - \theta^\circ_{\overline{m_-^\dagger}}\|^2_{l^2} \geqslant \tfrac{\kappa}{7}\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)\right)$$
$$+\ \tfrac{2}{7}\sum_{m\in[\![m_+^\dagger,n]\!]} \mathrm{pen}^\Lambda(m)\,\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant \tfrac{1}{7}\,\mathrm{pen}^\Lambda(m)\right) \quad (3.13)$$

The deviations of the last three terms in the last display (3.13) and also in (3.12) we bound by exploiting usual concentration inequalities which depend on the model considered. We hence formulate this property as the assumption to be verified in order to use this strategy.

Assumption 18 Remind that we defined $\Lambda_\circ(m) = \frac{1}{m}\sum_{s\in[\![1,m]\!]}\Lambda(s)$, $\Lambda_+(m) = \max\{\Lambda(s), s\in[\![1,m]\!]\}$, $\delta_\Lambda(m) \geqslant 1$ and $\Delta_\Lambda(m) = \delta_\Lambda(m)m\Lambda_+(m)$. Assume that there are numerical constants $(\mathcal{C}_i)_{1\in[\![1,11]\!]}$, such that for all $n\in\mathbb{N}$ and $m\in[\![1,n]\!]$ holds

(i) $\mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} - 12\frac{\Delta_\Lambda(m)}{n}\right)_+ \leqslant \mathcal{C}_1\Big[\frac{\mathcal{C}_2\,\Lambda_+(m)}{n}\exp\big(-\delta_\Lambda(m)m\mathcal{C}_3\big) + \frac{\mathcal{C}_4 m\Lambda_+(m)}{n^2}\exp\big(-\mathcal{C}_5\sqrt{n\delta_\Lambda(m)}\big)\Big]$

(ii) $\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\Delta_\Lambda(m)n^{-1}\right) \leqslant \mathcal{C}_6\Big[\exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big) + \exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big)\Big]$

(iii) $\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\mathfrak{R}_n^m(\theta^\circ,\Lambda)\right) \leqslant \mathcal{C}_9\Big[\exp\big(\frac{-\mathcal{C}_{10}n\mathfrak{R}_n^m(\theta^\circ,\Lambda)}{\Lambda_+(m)}\big) + \exp\big(\frac{-\mathcal{C}_{11}n\sqrt{\mathfrak{R}_n^m(\theta^\circ,\Lambda)}}{\sqrt{m\Lambda_+(m)}}\big)\Big]$

Consider now the partially data-driven aggregation of the orthogonal series estimators using either aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3) or model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.6) combining Assumption 18 and the upper bound given in (3.12) or (3.13) we obtain the next result, which proof is immediate and we omit it.

Lemma 3.2.5.
Assume that Assumption 18 holds true and use the penalty described in Definition 37 with

$\kappa \geqslant 84$ so that $\mathrm{pen}^{\Lambda}(m)/7 \geqslant 12n^{-1}\Delta_{\Lambda}(m)$for any m in $[\![1,n]\!]$. Then, for all $n \in \mathbb{N}$ and $m \in [\![1,n]\!]$ hold

(i) *let $m_{\mathcal{C}_3} := \lfloor 3(2/\mathcal{C}_3)^2 \rfloor$ and $n_{\mathcal{C}_5} := 15(\mathcal{C}_5)^{-4}$ then* $\sum_{m=1}^{n} \mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|_{l^2}^2 - \mathrm{pen}^{\Lambda}(m)/7\right)_+ \leqslant \mathcal{C}_1 n^{-1}\left[\frac{2\mathcal{C}_2}{\mathcal{C}_3}\Lambda_+(m_{\mathcal{C}_3}) + \mathcal{C}_4 n_{\mathcal{C}_5}\Lambda_+(n_{\mathcal{C}_5})\right]$

(ii) *let $m_{\mathcal{C}_7} := \lfloor 3(2/\mathcal{C}_7)^2 \rfloor$ and $n_{\mathcal{C}_8} := 15(3/\mathcal{C}_8)^4$ then*

$$\sum_{m=1}^{n} \mathrm{pen}^{\Lambda}(m)/7\,\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\Lambda}(m)/7\right) \leqslant \mathcal{C}_6 n^{-1}\left[\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_{\mathcal{C}_8})^2 n_{\mathcal{C}_8}^2\right]$$

(iii) $\mathbb{P}\left(\|\theta_{n,\overline{m^{\dagger}_-}} - \theta^{\circ}_{\overline{m^{\dagger}_n}}\|_{l^2}^2 \geqslant 12\mathfrak{R}_n^{m^{\dagger}_-}\right) \leqslant \mathcal{C}_9\left[\exp\left(\frac{-\mathcal{C}_{10} n \mathfrak{R}_n^{m^{\dagger}_-}}{\Lambda_+(m^{\dagger}_-)}\right) + (\mathcal{C}_8)^{-2} n^{-1}\right]$

□

Injecting Lemma 3.2.5 in either eq. (3.12) or eq. (3.13) we directly obtain the following result.

LEMMA 3.2.6.
Assume that Assumption 18 holds true. Consider the penalty sequence $\mathrm{pen}^{\Lambda}(m)$ as in Definition 37 with numerical constant $\kappa \geqslant 84$. Let $\widehat{\theta}^{(\eta)}$ be an aggregation estimator using either the aggregation weights eq. (3.3) or the model selection weights eq. (3.6). Let $n_{\mathcal{C}_5}$, $n_{\mathcal{C}_8}$, $m_{\mathcal{C}_3}$, and $m_{\mathcal{C}_7}$ be as in Lemma 3.2.5. Then, there is a finite numerical constant $\mathcal{C}$ such that for any $m^{\dagger}_-$, $m^{\dagger}_+$ and associated m_- and m_+ as in eq. (3.11) holds

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^{\circ}\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^{\Lambda}(m_+) + 2\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^{\circ}) + \mathcal{C}\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\left[\exp\left(-\mathcal{C}_{10}\delta_{\Lambda}(m^{\dagger}_-)m^{\dagger}_-\right)\right] + \mathcal{C}\left[\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2 \mathbb{1}_{\{m_->1\}} + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\right] n^{-1}. \quad (3.14)$$

□

The last bound allows us to derive an upper bound of the risk for data-driven aggregated estimator in the two cases **(p)** and **(np)** introduced in section 1.6.

THEOREM 3.2.1.
Under Assumption 18, consider the penalty sequence $\mathrm{pen}^{\Lambda}(m) := \kappa\,\Delta_{\Lambda}(m)\,n^{-1}$, $m \in [\![1,n]\!]$, as in Definition 37 with numerical constant $\kappa \geqslant 84$. Let $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_M^{(\eta)}(m)\theta_{n,\overline{m}}$ be an aggregation of the orthogonal series estimators, using either aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3), or model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.6).

(p) Assume there is $K \in \mathbb{N}$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^{\circ}) > 0$ and $\mathfrak{b}_m(\theta^{\circ}) = 0$. For $K > 0$ let $c_{\theta^{\circ}} := \frac{\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2 + 4\kappa}{\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^{\circ})} > 1$ and $n_{\theta^{\circ}} := \lfloor c_{\theta^{\circ}}\Delta_{\Lambda}(K)\rfloor \in \mathbb{N}$. If $n \in [\![1,n_{\theta^{\circ}}]\!]$ then set $m^{\bullet}_n := m_{\mathcal{C}_3}\log(n)$, and otherwise if $n > n_{\theta^{\circ}}$ then set $m^{\bullet}_n := \max\{m \in [\![K,n]\!] : n > c_{\theta^{\circ}}\Delta_{\Lambda}(m)\}$ where the defining set contains K and thus is not empty. There is a finite constant $\mathcal{C}_{\theta^{\circ},\Lambda}$ given in (B.23) depending only on θ° and Λ such that for all $n \in \mathbb{N}$ holds

$$\mathcal{R}_n\left(\widehat{\theta}^{(\eta)}, \theta^{\circ}, \Lambda\right) \leqslant \mathcal{C}\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\left[n^{-1} \vee \exp\left(-\mathcal{C}_{10}\delta_{\Lambda}(m^{\bullet}_n)m^{\bullet}_n\right)\right] + \mathcal{C}_{\theta^{\circ},\Lambda} n^{-1}. \quad (3.15)$$

(np) *Assume that* $\mathfrak{b}_m(\theta^\circ) > 0$ *for all* $m \in \mathbb{N}$. *There is a finite finite constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *given in* (B.24) *depending only* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}(\|\theta^\circ_{\underline{0}}\|^2_{l^2} \vee 1) \min_{m\in[\![1,n]\!]} \big[\mathcal{R}^m_n(\theta^\circ, \Lambda) \vee \exp\big(-\mathcal{C}_{10}\delta_\Lambda(m)m\big)\big] + \mathcal{C}_{\theta^\circ,\Lambda} n^{-1}. \tag{3.16}$$

Hence, using Theorem 3.2.1 gives us the following result.

Corollary 3.2.1.
Let the assumptions of Theorem 3.2.1 be satisfied.

(p) *If in addition (A1) there is* $n_{\theta^\circ,\Lambda} \in \mathbb{N}$ *such that* $\delta_\Lambda(m^\bullet_n)m^\bullet_n \geqslant (\mathcal{C}_{10})^{-1}(\log n)$ *for all* $n \geqslant n_{\theta^\circ,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *depending only on* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda} n^{-1}$.

(np) *If in addition (A2) there is* $n_{\theta^\circ,\Lambda} \in \mathbb{N}$ *such that*

$m^\dagger_n(\theta^\circ)\delta_\Lambda(m^\dagger_n(\theta^\circ)) \geqslant (\mathcal{C}_{10})^{-1}|\log \mathfrak{R}^\dagger_n(\theta^\circ, \Lambda)|$ *for all* $n \geqslant n_{\theta^\circ,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *depending only on* θ° *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\mathfrak{R}^\dagger_n(\theta^\circ, \Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.*

Numerical discussion 3.2.2.
Let us briefly illustrate the last results. In case (p) the partially data-driven aggregation leads to an estimator attaining the parametric oracle rate (see Remark 1.3.1), if the additional assumption (A1) is satisfied. Consider the two cases **(o)** and **(s)** for λ as in Num. discussion 1.3.1:

(o) $1 \approx \Delta_\Lambda(m^\bullet_n)n^{-1} \approx (m^\bullet_n)^{2a+1}n^{-1}$ implies $m^\bullet_n \approx n^{1/(2a+1)}$ and $m^\bullet_n\delta_\Lambda(m^\bullet_n) \approx n^{1/(2a+1)}$

(s) $n \approx \Delta_\Lambda(m^\bullet_n) \approx (m^\bullet_n)^{1+4a}\exp((m^\bullet_n)^{2a})$ implies $m^\bullet_n \approx (\log n - \frac{1+4a}{2a}\log\log n)^{1/(2a)}$ and $m^\bullet_n\delta_\Lambda(m^\bullet_n) \approx (\log n)^{2+1/(2a)}$.

Clearly in both cases **(o)** and **(s)**, the additional condition (A1) of corollary 3.2.1 holds true. Therefore, in this situation the aggregated estimator attains the oracle rate. On the other hand side, in case (np) the partially data-driven aggregation leads to an estimator attaining the rate $\mathfrak{R}^\dagger_n(\theta^\circ, \Lambda)$ (see Remark 1.3.1), if the additional assumption (A2) is satisfied. Otherwise, the upper bound faces a deterioration of the rate, which we illustrate considering as in Num. discussion 3.2.1 usual behaviour **[o-o]**, **[o-s]** and **[s-o]** for the sequences $(\mathfrak{b}_m(\theta^\circ))_{m\in\mathbb{N}}$ and $(\Lambda(m))_{m\in\mathbb{N}}$. In case **[o-o]**, **[o-s]** and **[s-o]** only with $p < 1/2$ the assumption (A2) is satisfied, and $\mathfrak{R}^\dagger_n(\theta^\circ, \Lambda)$ equals the oracle rate $\mathcal{R}^\circ_n(\theta^\circ, \Lambda)$ (cf. **[o-o]** **[o-o]**, **[o-s]** and **[s-o]**). Thereby, the partially data-driven aggregation leads to an estimator attaining the oracle rate $\mathcal{R}^\circ_n(\theta^\circ, \Lambda)$. In case **[o-s]** with $p \geqslant 1/2$ the assumption (A2) is not satisfied. However, with $m^\bullet_n := m_{\mathcal{C}_3}|\log \mathfrak{R}^\dagger_n| \approx (\log n)$ holds $\min_{m\in[\![1,n]\!]}\big[\mathcal{R}^m_n(\theta^\circ, \Lambda) \vee \exp\big(\frac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_3}}\big) \leqslant \mathfrak{R}^{m^\bullet_n}_n\theta^\circ, \Lambda \approx (\log n)^{2a+1}n^{-1}$. In this situation the rate of the partially data-driven estimator $\widehat{\theta}^{(\eta)}$ features a deterioration by a logarithmic factor $(\log n)^{(2a+1)(1-1/(2p))}$ compared to the oracle rate, i.e. $\mathfrak{R}^{m^\bullet_n}_n \approx (\log n)^{2a+1}n^{-1}$ versus $\mathcal{R}^\circ_n \approx (\log n)^{(2a+1)/(2p)}n^{-1}$. □

3.2.1.2 Maximal risk bounds

By applying Lemma 3.2.2, we derive bounds for the maximal risk over ellipsoids $\Theta(\mathfrak{a}, r)$ of the aggregated estimator $\widehat{\theta}^{(\eta)}$ using either aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3) or model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.4). Therefore, we aim next to control the second and third right hand side term in (3.9) uniformly over $\Theta(\mathfrak{a}, r)$. Keeping the definition (1.6) of $\mathcal{R}_n^m(\mathfrak{a}, \Lambda)$ in mind it holds $r^2\mathcal{R}_n^m(\mathfrak{a}, \Lambda) \geqslant \|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ)$ uniformly for all $\theta^\circ \in \Theta(\mathfrak{a}, r)$ and for all $m \in \mathbb{N}$. The proofs for the results displayed here can be found in appendix B.1.2. We then gives the following definition for the sequence which we want to prove to be an upper bound for the maximal risk of the aggregation estimator. Note that in this case we use $\Delta_\Lambda(m)$ and $\mathrm{pen}^\Lambda(m)$ as defined in Definition 37 and hence the rates for the quadratic as well as the maximal risk are obtained for the same estimator.

Definition 40 Let be the following family of sequences, $\mathfrak{R}_n^m(\mathfrak{a}) := \mathfrak{R}_n^m(\mathfrak{a}, \Lambda) := [\mathfrak{a}(m)^2 \vee \Delta_\Lambda(m)\, n^{-1}]$. Then it holds for all m in $[\![1, n]\!]$ and θ° in $\Theta(\mathfrak{a}, r)$

$$[r^2 + \kappa]\mathfrak{R}_n^m(\mathfrak{a}) \geqslant \left[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ) \vee \mathrm{pen}^\Lambda(m)\right] \tag{3.17}$$

Considering the following specific case, we aim to show that it describes an upper bound for the maximal risk over $\Theta(\mathfrak{a}, r)$ for our aggregation estimator,

$$m_n^\dagger(\mathfrak{a}) := \arg\min\left\{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\right\} \in [\![1, n]\!]$$
$$\mathfrak{R}_n^\dagger(\mathfrak{a}) := \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) := \min\left\{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\right\};\ \text{with } \mathfrak{R}_n^{m_n^\dagger(\mathfrak{a})}(\mathfrak{a}, \Lambda) = \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)$$

□

Numerical discussion 3.2.3.
Let us illustrate Definition 40 considering as in Num. discussion 1.3.4 usual behaviour **[o-o]**, **[s-o]** and **[o-s]** for the sequences $(\mathfrak{a}(m))_{m\in\mathbb{N}}$ and $(\Lambda(m))_{m\in\mathbb{N}}$:

[o-o] Since $\Delta_\Lambda(m) \approx m^{2a+1}$ follows $m_n^\dagger(\mathfrak{a}) \approx n^{1/(2p+2a+1)}$, $\delta_\Lambda(m_n^\dagger(\mathfrak{a}))m_n^\dagger(\mathfrak{a}) \approx n^{1/(2p+2a+1)}$, $\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) \approx n^{-2p/(2p+2a+1)}$ and $|\log \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)| \approx (\log n)$.

[o-s] Since $\Delta_\Lambda(m) \approx m^{1+4a}\exp(m^{2a})$ follows $m_n^\dagger(\mathfrak{a}) \approx (\log n)^{1/(2a)}$, $\delta_\Lambda(m_n^\dagger(\mathfrak{a}))m_n^\dagger(\mathfrak{a}) \approx (\log n)^{2+1/(2a)}$, $\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) \approx (\log n)^{-p/a}$ and $|\log \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)| \approx (\log\log n)$.

[s-o] Since $\Delta_\Lambda(m) \approx m^{2a+1}$ follows $m_n^\dagger(\mathfrak{a}) \approx (\log n)^{1/(2p)}$, $\delta_\Lambda(m_n^\dagger(\mathfrak{a}))m_n^\dagger(\mathfrak{a}) \approx (\log n)^{1/(2p)}$, $\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) \approx (\log n)^{(2a+1)/(2p)}n^{-1}$ and $|\log \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)| \approx (\log n)$. □

We note that in the three cases **[o-o]**, **[o-s]** and **[s-o]** the rate $\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)$ coincide with the minimax rate $\mathcal{R}_n^\star(\mathfrak{a}, \Lambda)$ derived in Num. discussion 1.3.4 **[o-o]**, **[o-s]** and **[s-o]**, respectively.

□

Keeping in mind (3.17) for any $m_+^\dagger, m_-^\dagger \in [\![1, n]\!]$ let us define

$$m_- := \min\left\{m \in [\![1, m_-^\dagger]\!] : \|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ) \leqslant [r^2 + 4\kappa]\mathcal{R}_n^{m_-^\dagger}(\mathfrak{a})\right\} \quad \text{and}$$
$$m_+ := \max\left\{m \in [\![m_+^\dagger, n]\!] : \mathrm{pen}^\Lambda(m) \leqslant 2[3r^2 + 2\kappa]\mathcal{R}_n^{m_+^\dagger}(\mathfrak{a})\right\} \tag{3.18}$$

where the defining sets obviously contains $m_-^\dagger$ and $m_+^\dagger$, repsectively, and hence, they are not empty.

Lemma 3.2.7.
Consider the data-driven aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3) and the rates described in Definition 40 with $\kappa \geqslant 8\log(3e)$ for any $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ and associated $m_+, m_- \in [\![1,n]\!]$ as in (3.18) hold

(i) $\mathbb{P}_M^{(\eta)}([\![1,m_-[\![)\mathbb{1}_{\left\{\|\theta_{\overline{n,m_-^\dagger}}-\theta^\circ_{\overline{m_-^\dagger}}\|_{l^2}^2<\kappa\mathfrak{R}_n^{m_-^\dagger}(\mathfrak{a})/7\right\}} \leqslant \frac{1}{\eta\kappa}\mathbb{1}_{\{m_->1\}}\exp\left(-\frac{3\eta\kappa}{14}n\mathfrak{R}_n^{m_-^\dagger}(\mathfrak{a})\right);$

(ii) $\sum_{m\in]\!]m_+,n]\!]}\mathrm{pen}^\Lambda(m)\,\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2<\mathrm{pen}^\Lambda(m)/7\}} \leqslant n^{-1}\{\frac{16}{\kappa\eta^2}+\frac{8}{\eta}\}.$

□

The next result can also immediately be deduced from Lemma 3.2.7 letting $\eta\to\infty$. On the other hand side, a direct proof follows line by line the proof of Lemma 3.2.4 using (3.17) rather than (3.10), and we omit the details.

Lemma 3.2.8.
Consider the data-driven model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.4). Under definition Definition 40 for any $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ and associated $m_+, m_- \in [\![1,n]\!]$ as in (3.18) hold

(i) $\mathbb{P}_M^{(\infty)}([\![1,m_-[\![)\mathbb{1}_{\{\|\theta_{\overline{n,m_-^\dagger}}-\theta^\circ_{\overline{m_-^\dagger}}\|_{l^2}^2<\kappa\mathfrak{R}_n^{m_-^\dagger}(\mathfrak{a})/7\}} = 0;$

(ii) $\sum_{m\in]\!]m_+,n]\!]}\mathrm{pen}^\Lambda(m)\,\mathbb{P}_M^{(\infty)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2<\mathrm{pen}^\Lambda(m)/7\}} = 0.$

□

Lemma 3.2.9.
Assume that Assumption 18 holds true and use the penalty described in Definition 37 with $\kappa \geqslant 84$ so that $\mathrm{pen}^\Lambda(m)/7 \geqslant 12n^{-1}\Delta_\Lambda(m)$for any m in $[\![1,n]\!]$. Then, for all $n\in\mathbb{N}$ and $m\in[\![1,n]\!]$ hold

(i) *let* $m_{\mathcal{C}_3} := \lfloor 3(2/\mathcal{C}_3)^2\rfloor$ *and* $n_{\mathcal{C}_5} := 15(\mathcal{C}_5)^{-4}$ *then*
$$\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\sum_{m=1}^n \mathbb{E}\left(\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2-\mathrm{pen}^\Lambda(m)/7\right)_+ \leqslant \mathcal{C}_1 n^{-1}\left[\Lambda_+(m_{\mathcal{C}_3})+\Lambda_+(n_{\mathcal{C}_5})\right]$$

(ii) *let* $m_{\mathcal{C}_7} := \lfloor 3(2/\mathcal{C}_7)^2\rfloor$ *and* $n_{\mathcal{C}_8} := 15(3/\mathcal{C}_8)^4$ *then*
$$\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\sum_{m=1}^n \frac{\mathrm{pen}^\Lambda(m)}{7}\,\mathbb{P}\left(\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|_{l^2}^2 \geqslant \frac{\mathrm{pen}^\Lambda(m)}{7}\right) \leqslant \mathcal{C}_6 n^{-1}\left[\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2+\Lambda_+(n_{\mathcal{C}_8})^2\right]$$

(iii) $$\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\mathbb{P}\left(\|\theta_{\overline{n,m_-^\dagger}}-\theta^\circ_{\overline{m_n^\dagger}}\|_{l^2}^2 \geqslant 12\mathfrak{R}_n^{m_-^\dagger}\right) \leqslant \mathcal{C}_9\left[\exp\left(\frac{-\mathcal{C}_{10}n\mathfrak{R}_n^{m_-^\dagger}}{\Lambda_+(m_-^\dagger)}\right)+n^{-1}\right]$$

□

Consider now the partially data-driven aggregation of the orthogonal series estimators using either aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3) or model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.4). From (3.9), combining Lemma 3.2.7 and Lemma 3.2.8 we obtain by replacing $\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ)$ by $\mathfrak{R}_n^{m_-^\dagger}(\mathfrak{a})$ upper bounds similar to (3.12) and (3.13), respectively. Those upper bounds together with Lemma 3.2.9 allow us to show the next assertion Lemma 3.2.10.

Lemma 3.2.10.
Assume that Assumption 18 holds true.
Consider the penalty sequence $\mathrm{pen}^{\Lambda}(m) := \kappa\, \Delta_{\Lambda}(m)\, n^{-1}$, $m \in [\![1, n]\!]$, *as in Definition 37 with numerical constant* κ *to be specified depending on the model.*
Let $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_{M}^{(\eta)}(m)\theta_{n,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\mathbb{P}_{M}^{(\eta)}$ *as in (3.3) or model selection weights* $\mathbb{P}_{M}^{(\infty)}$ *as in (3.4). There is a finite numerical constant* $\mathcal{C} > 0$ *such that for any* $\theta^{\circ} \in \Theta(\mathfrak{a}, r)$, $m_{-}^{\dagger}, m_{+}^{\dagger} \in [\![1, n]\!]$ *and associated* $m_{+}, m_{-} \in [\![1, n]\!]$ *as defined in (3.18) hold*

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^{\circ}\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^{\Lambda}(m_{+}) + 2\|\theta_{\underline{0}}^{\circ}\|_{l^2}^2\, \mathfrak{b}_{m_{-}}^2(\theta^{\circ}) + \mathcal{C}\mathfrak{R}_n^{\dagger}(\mathfrak{a}). \tag{3.19}$$

□

The last bound allows us to derive an upper bound of the maximal risk over the ellipsoid $\Theta(\mathfrak{a}, r)$ for the partially data-driven aggregated estimator in the case **(np)** introduced in section 1.6.

Theorem 3.2.2.
Assume that Assumption 18 holds true and consider the penalty sequence $\mathrm{pen}^{\Lambda}(m) := \kappa\, \Delta_{\Lambda}(m)\, n^{-1}$, $m \in [\![1, n]\!]$, *as in Definition 37. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_{M}^{(\eta)}(m)\theta_{n,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\mathbb{P}_{M}^{(\eta)}$ *as in (3.3) or model selection weights* $\mathbb{P}_{M}^{(\infty)}$ *as in (3.4). There is a finite constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *given in (B.24) depending only on* $\mathfrak{a}$, r *and* Λ *such that for all* $n \in \mathbb{N}$ *and for all* $m_n^{\bullet} \in [\![m_n^{\dagger}(\mathfrak{a}), n]\!]$ *with* $m_n^{\dagger}(\mathfrak{a}) \in [\![1, n]\!]$ *as in Definition 40 holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}(r^2 \vee 1) \min_{m \in [\![1,n]\!]} \big[\mathfrak{R}_n^{m}(\mathfrak{a}, \Lambda) \vee \exp\big(- \mathcal{C}_{10}\delta_{\Lambda}(m)m)\big]\big] + \mathcal{C}_{\mathfrak{a},r,\Lambda}n^{-1}. \tag{3.20}$$

□

Corollary 3.2.2.
Let the assumptions of Theorem 3.2.2 be satisfied. If in addition **(A)** *there is* $n_{\mathfrak{a},r,\Lambda} \in \mathbb{N}$ *such that* $m_n^{\dagger}(\mathfrak{a})\delta_{\Lambda}(m_n^{\dagger}(\mathfrak{a})) \geqslant (\mathcal{C}_{10})^{-1}|\log \mathfrak{R}_n^{\dagger}(\mathfrak{a})|$ *for all* $n \geqslant n_{\mathfrak{a},r,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\Theta(\mathfrak{a}, r)$ *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\mathfrak{R}_n^{\dagger}(\mathfrak{a}, \Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.*

Numerical discussion 3.2.4.
Let us illustrate Theorem 3.2.2 and corollary 3.2.2. Under corollary 3.2.2 the partially data-driven aggregated estimator attains the rate $\mathfrak{R}_n^{\dagger}(\mathfrak{a}, \Lambda)$. Otherwise, the upper bound faces a deterioration of the rate, which we illustrate considering as in Num. discussion 1.3.4 usual behaviour **[o-o]**, **[o-s]** and **[s-o]** for the sequences $(\mathfrak{a}(m))_{m\in\mathbb{N}}$ and $(\Lambda(m))_{m\in\mathbb{N}}$. In case **[o-o]**, **[o-s]** and **[s-o]** only with $p < 1/2$ the assumption **(A)** is satisfied, and $\mathfrak{R}_n^{\dagger}(\mathfrak{a}, \Lambda)$ equals the oracle rate $\mathcal{R}_n^{\circ}(\mathfrak{a}, \Lambda)$ (cf. Num. discussion 3.2.3 **[o-o]**, **[o-s]** and **[s-o]**). Thereby, the partially data-driven aggregation leads to an estimator attaining the oracle rate $\mathcal{R}_n^{\circ}(\mathfrak{a}, \Lambda)$. In case **[o-s]** with $p \geqslant 1/2$ the assumption **(A)** is not satisfied. However, with $m_n^{\bullet} := m_{\mathcal{C}_3}|\log \mathfrak{R}_n^{\dagger}(\mathfrak{a}, \Lambda)| \approx (\log n)$ holds $\min_{m\in[\![1,n]\!]} \big[\mathcal{R}_n^{m}(\mathfrak{a}, \Lambda) \vee \exp\big(\frac{-\delta_{\Lambda}(m)m}{m_{\mathcal{C}_3}}\big) \leqslant \mathfrak{R}_n^{m_n^{\bullet}}(\mathfrak{a}, \Lambda) \approx$

$(\log n)^{2a+1}n^{-1}$. In this situation the rate of the partially data-driven estimator $\widehat{\theta}^{(\eta)}$ features a deterioration by a logarithmic factor $(\log n)^{(2a+1)(1-1/(2p))}$ compared to the oracle rate, i.e. $\mathfrak{R}_n^{m_n^{\bullet}}(\mathfrak{a},\Lambda) \approx (\log n)^{2a+1}n^{-1}$ versus $\mathcal{R}_n^{\circ}(\mathfrak{a},\Lambda) \approx (\log n)^{(2a+1)/(2p)}n^{-1}$. □

3.2.2 Unknown operator

Consider now the case Assumption 4. We shall hence keep in mind 3.2, 3.4, Definition 38 as well as 3.7 and 3.8. **Note that the detailed proofs for all results given here can be found in appendix B.2.**

We will assume, from now on, that Assumption 9 holds true.

Both for the quadratic and the maximal risk, our strategy is based on the decomposition of the quadratic loss function displayed in Lemma 3.2.2. This decomposition is independent of the model an only relies on the fact that the parameter space is equipped with a nested sieve and the fact that our estimator aggregation structure takes advantage of it.

Lemma 3.2.11.
Consider the aggregated OSE $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ *with weights* $\mathbb{P}_M^{(\eta)}(m) \in [0,1]$, $m \in [\![1,n]\!]$, *satisfying* $\sum_{m=1}^{n} \mathbb{P}_M^{(\eta)}(m) = 1$ *and a sequence* $(\mathrm{pen}(m))_{m\in[\![1,n]\!]}$ *of non-negative compensation terms. Given* $m \in \mathbb{N}$ *let* $\breve{\theta}_{\overline{m}} := \sum_{s=-m}^{m} \lambda_{n_\lambda}^{+}(s)\phi(s)$. *For any* $m_- \in [\![1,n]\!]$, $m_+ \in [\![1,n]\!]$, *and the sequence of events* $(\mathcal{X}_s)_{s\in\mathbb{Z}} = (\{|\lambda_{n_\lambda}^{+}(s)|^2 \geqslant n_\lambda^{-1}\})_{s\in\mathbb{Z}}$ *holds*

$$\begin{aligned}
\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 3\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|_{l^2}^2 + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_-}^2(\theta^\circ) \\
+ 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathbb{P}_M^{(\eta)}([\![1,m_-[\![) + \tfrac{3}{7}\sum_{l\in]\!]m_+,n]\!]} \mathrm{pen}(l)\,\mathbb{P}_M^{(\eta)}(l)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{l}}-\breve{\theta}_{\overline{l}}\|_{l^2}^2<\mathrm{pen}(l)\}} \\
+ 3\sum_{l\in]\!]m_+,n]\!]} \left(\|\theta_{n,n_\lambda,\overline{l}} - \breve{\theta}_{\overline{l}}\|_{l^2}^2 - \mathrm{pen}(l)/7\right)_+ + \tfrac{3}{7}\sum_{l\in]\!]m_+,n]\!]} \mathrm{pen}(l)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{l}}-\breve{\theta}_{\overline{l}}\|_{l^2}^2\geqslant\mathrm{pen}(l)/7\}} \\
+ 6\sum_{s\in[\![1,n]\!]} |\lambda_{n_\lambda}^{+}(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\theta^\circ(s)|^2 + 2\sum_{s\in[\![1,n]\!]} \mathbb{1}_{\mathcal{X}_s^c}|\theta^\circ(s)|^2 \quad (3.21)
\end{aligned}$$

Keep in mind the shape of the estimator given in 3.2, eq. (3.4), eq. (3.7), eq. (3.8), and Definition 38.

3.2.2.1 Quadratic risk bounds

We derive bounds for the risk of the aggregated estimator $\widehat{\theta}^{(\eta)}$ and the model selected estimator $\theta_{n,\widehat{\overline{m}}}$ by applying Lemma 3.2.11. Therefore, we aim next to control the third and fourth right hand side term in (3.21). The proofs for the results stated here can be found in appendix B.2.1.

For each $m \in \mathbb{N}$ keep in mind that $\|\theta_{\underline{m}}^\circ\|_{l^2}^2 = \|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ)$, $\mathfrak{R}_n^m(\theta^\circ,\Lambda) := [\mathfrak{b}_m^2(\theta^\circ) \vee \Delta_\Lambda(m)\, n^{-1}]$ as in Definition 40 and introduce in addition $\breve{\theta}_{\overline{m}} = \mathbb{1}_{\{|s|\leqslant m\}}\lambda_{n_\lambda}^{+}(s)\phi(s)$. Note that $\breve{\theta}_{\overline{m}} = \Pi_{\overline{m}}\breve{\theta}_{\overline{n}}$ and $\|\Pi_{\overline{m}}^{\perp}\breve{\theta}_{\overline{n}}\|_{l^2}^2 = 2\sum_{s\in]\!]m,n]\!]} \widehat{\Lambda}(s)|\phi(s)|^2$. For any $m_+^\dagger, m_-^\dagger \in [\![1,n]\!]$ let us define

$$m_- := \min\Big\{m \in [\![1, m_-^\dagger]\!] : \|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ) \leqslant [\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa]\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda)\Big\} \quad \text{and}$$

$$m_+ := \max\Big\{m \in [\![m_+^\dagger, n]\!] : \mathrm{pen}^{\widehat{\Lambda}}(m) \leqslant 2[3\|\Pi_{\overline{m_+^\dagger}}^{\perp}\breve{\theta}_{\overline{n}}\|_{l^2}^2 + 2\,\mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger)]\Big\} \tag{3.22}$$

where the defining set obviously contains $m_-^\dagger$ and $m_+^\dagger$, respectively, and hence, they are not empty. Keep in mind that $m_+ := m_+(\varepsilon_1, \dots, \varepsilon_{n_\lambda})$ is random but does not depend on the sample $Y_1, \dots, Y_n$.

Lemma 3.2.12.
Consider the data-driven aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4). Using the penalty as in Definition 38 with $\mho_l := \{1/4 \leqslant \Lambda(s)^{-1}\widehat{\Lambda}(s) \leqslant 9/4,\ \forall\, s \in [\![1, l]\!]\}$, $l \in [\![1, n]\!]$, for any $m_-^\dagger, m_+^\dagger \in [\![1, n]\!]$ and associated $m_+, m_- \in [\![1, n]\!]$ as in (3.22) hold

(i) $\mathbb{P}_M^{(\eta)}([\![1, m_-[\![)$
$\leqslant \frac{50}{\eta\kappa}\mathbb{1}_{\{m_->1\}} \exp\big(-\frac{\eta\kappa}{2} n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda)\big) + \mathbb{1}_{\{\|\theta_{\overline{n,n_\lambda,m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\} \cup \mho_{m_-^\dagger}^c}$;

(ii) $\sum_{m \in]\!]m_+, n]\!]} \mathrm{pen}^{\widehat{\Lambda}}(m)\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 < \mathrm{pen}^{\widehat{\Lambda}}(m)/7\}} \leqslant n^{-1}\{\frac{16}{\kappa\eta^2} + \frac{8}{\eta}\}$.

□

We combine the upper bound in (3.21) and the bounds given in Lemma 3.2.12. Conditionally on $\varepsilon_1, \dots, \varepsilon_{n_\lambda}$ the r.v.'s $Y_1, \dots, Y_n$ are iid. and we denote by $\mathbb{P}_{Y|\varepsilon}$ and $\mathbb{E}_{Y|\varepsilon}$ their conditional distribution and expectation, respectively. Clearly, due to Lemma 3.2.12 we have

$$\begin{aligned}\mathbb{E}_{Y|\varepsilon}\,\widehat{\mathbb{P}}_M^{(\eta)}([\![1, m_-[\![) \leqslant \mathbb{1}_{\{m_->1\}}\big[&\tfrac{50}{\eta\kappa}\exp\big(-\tfrac{3\eta\kappa}{14} n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda)\big)\\ &+ \mathbb{P}_{Y|\varepsilon}\big(\|\theta_{\overline{n,n_\lambda,m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\big)\mathbb{1}_{\mho_{m_-^\dagger}} + \mathbb{1}_{\mho_{m_-^\dagger}^c}\big]\end{aligned}$$

and, hence from (3.21) follows immediately

$$\begin{aligned}\mathbb{E}_{Y|\varepsilon}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant{}& 3\,\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|_{l^2}^2 + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ &+ \tfrac{150}{\eta\kappa}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(-\tfrac{3\eta\kappa}{14} n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda)\big) + \tfrac{3}{7}n^{-1}\{\tfrac{16}{\kappa\eta^2} + \tfrac{8}{\eta}\}\\ &+ 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\big[\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{\overline{n,n_\lambda,m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\big)\mathbb{1}_{\mho_{m_-^\dagger}} + \mathbb{1}_{\mho_{m_-^\dagger}^c}\big]\\ &+ 3\sum_{l \in]\!]m_+, n]\!]}\mathbb{E}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{l}} - \breve{\theta}_{\overline{l}}\|_{l^2}^2 - \mathrm{pen}(l)/7\big)_+\\ &+ \tfrac{3}{7}\sum_{l \in]\!]m_+, n]\!]}\mathrm{pen}^{\widehat{\Lambda}}(l)\,\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{l}} - \breve{\theta}_{\overline{l}}\|_{l^2}^2 \geqslant \mathrm{pen}(l)/7\big)\\ &+ 6\sum_{s \in [\![1, n]\!]}|\lambda_{n_\lambda}^+(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\theta^\circ(s)|^2 + 2\sum_{s \in [\![1, n]\!]}\mathbb{1}_{\mathcal{X}_s^c}|\theta^\circ(s)|^2\end{aligned} \tag{3.23}$$

The next result can be directly deduced from Lemma 3.2.12 by letting $\eta \to \infty$. However, we think the direct proof given in annex provides an interesting illustration of the values $m_+, m_- \in [\![1,n]\!]$ as defined in (3.22).

Lemma 3.2.13.
Consider the data-driven model selection weights $\mathbb{P}_M^{(\infty)}$ as in eq. (3.8). Under definition Definition 38 for any $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ and associated $m_+, m_- \in [\![1,n]\!]$ as in (3.22) hold

(i) $\mathbb{P}_M^{(\infty)}([\![1, m_-[\![)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 < \widehat{\mathrm{pen}^\Lambda}(m_-^\dagger)/7\}\cap\mho_{m_-^\dagger}} = 0;$

(ii) $\sum_{m\in]\!]m_+,n]\!]} \mathrm{pen}^{\widehat{\Lambda}}(m)\, \mathbb{P}_M^{(\infty)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 < \mathrm{pen}^{\widehat{\Lambda}}(m)/7\}} = 0.$

□

We combine the upper bound in (3.21) and the bounds given in Lemma 3.2.13. Clearly, due to Lemma 3.2.13 we have

$$\mathbb{E}_{Y|\varepsilon}\, \mathbb{P}_M^{(\infty)}([\![1, m_-[\![) \leqslant \mathbb{1}_{\{m_->1\}}\big[\, \mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\big)\mathbb{1}_{\mho_{m_-^\dagger}} + \mathbb{1}_{\mho_{m_-^\dagger}^c}\,\big]$$

and, hence from (3.21) follows immediately

$$\begin{aligned}
\mathbb{E}_{Y|\varepsilon}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 3\,\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|_{l^2}^2 + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_-}^2(\theta^\circ) \\
+ 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\big[\, \mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\big)\mathbb{1}_{\mho_{m_-^\dagger}} + \mathbb{1}_{\mho_{m_-^\dagger}^c}\,\big] \\
+ 3\sum_{l\in]\!]m_+,n]\!]} \mathbb{E}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\bar{l}} - \breve{\theta}_{\bar{l}}\|_{l^2}^2 - \mathrm{pen}(l)/7\big)_+ \\
+ \tfrac{3}{7}\sum_{l\in]\!]m_+,n]\!]} \mathrm{pen}^{\widehat{\Lambda}}(l)\, \mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\bar{l}} - \breve{\theta}_{\bar{l}}\|_{l^2}^2 \geqslant \mathrm{pen}(l)/7\big) \\
+ 6\sum_{s\in[\![1,n]\!]} |\lambda_{n_\lambda}^+(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\theta^\circ(s)|^2 + 2\sum_{s\in[\![1,n]\!]} \mathbb{1}_{\mathcal{X}_s^c}|\theta^\circ(s)|^2 \quad (3.24)
\end{aligned}$$

The deviations of the last three terms in the last display (3.24) and also in (3.23) need to be bounded using concentration inequalities which depend on the considered model. We hence formulate it as the central hypothesis to be verified in order to apply this method.

Assumption 19 Consider $\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}} = \sum_{|s|\in[\![1,m]\!]} \lambda_{n_\lambda}^+(s)(\phi_n(s) - \phi(s))e_s$. Conditionally on $\{\varepsilon_1, \ldots, \varepsilon_{n_\lambda}\}$ the r.v.'s $\{Y_1, \ldots, Y_n\}$ are iid. and we denote by $\mathbb{P}_{Y|\varepsilon}$ and $\mathbb{E}_{Y|\varepsilon}$ their conditional distribution and expectation, respectively. Let $\widehat{\Lambda}(s) = |\lambda_{n_\lambda}^+(s)|^2$, $\Lambda_\circ(m) = \frac{1}{m}\sum_{s\in[\![1,m]\!]} \widehat{\Lambda}(s)$, $\widehat{\Lambda}_+(m) = \max\{\widehat{\Lambda}(s), s \in [\![1,m]\!]\}$, $\Delta_{\widehat{\Lambda}}(m) = \delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)$ and $\delta_{\widehat{\Lambda}}(m) \geqslant 1$. Then there is a numerical constant $\mathcal{C}$ such that for all $n \in \mathbb{N}$ and $m \in [\![1,n]\!]$ holds

(i) $\mathbb{E}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 - 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big)_+$
$\leqslant \mathcal{C}_1\Big[\frac{\mathcal{C}_2\widehat{\Lambda}_+(m)}{n}\exp\big(-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m\big) + \frac{\mathcal{C}_4 m\widehat{\Lambda}_+(m)}{n^2}\exp\big(-\mathcal{C}_5\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big)\Big]$

(ii) $\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big)$
$\leqslant \mathcal{C}_6\Big[\exp\big(-\mathcal{C}_7\delta_{\widehat{\Lambda}}(m)m\big) + \exp\big(-\mathcal{C}_8\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big)\Big]$

(iii) $\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big)$
$\leqslant \mathcal{C}_9\Big[\exp\big(-\mathcal{C}_{10}\delta_{\widehat{\Lambda}}(m)m\big) + \exp\big(\frac{-\mathcal{C}_{11}n\sqrt{\mathfrak{R}_n^m\theta^\circ,\widehat{\Lambda}}}{\sqrt{m\widehat{\Lambda}_+(m)}}\big)\Big]$

(iv) $\mathbb{P}\big(|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\big) \leqslant \mathcal{C}_{12}\exp\big(-\mathcal{C}_{13}n_\lambda|\lambda(s)|^2\big) \leqslant \mathcal{C}_{14}\exp\big(-\frac{\mathcal{C}_{15}n_\lambda}{\Lambda_+(m)}\big)$.

□

This hypothesis allows us to control the remaining random elements in our bound.

Lemma 3.2.14.
Consider $\mathrm{pen}^{\widehat{\Lambda}}(m) = \kappa\,\Delta_{\widehat{\Lambda}}(m)\,n^{-1}$, $m \in [\![1,n]\!]$, *as in Definition 38 with* $\kappa \geqslant 84$. *Let* $m_{\mathcal{C}_3} := [\lfloor 3(\frac{2}{\mathcal{C}_3})^2\rfloor \vee \mathcal{C}_2]$ *and* $n_{\mathcal{C}_5} := 15(\frac{1}{\mathcal{C}_5})^4$*; as well as* $m_{\mathcal{C}_7} := \lfloor 3(\frac{2}{\mathcal{C}_7})^2\rfloor$ *and* $n_{\mathcal{C}_8} := \lfloor 15(3/\mathcal{C}_8)^4\rfloor$. *There exists a finite numerical constant* $\mathcal{C} > 0$ *such that for all* $n \in \mathbb{N}$ *and all* $m_-^\dagger \in [\![1,n]\!]$ *hold*

(i) $\sum_{m=1}^n \mathbb{E}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 - \mathrm{pen}^{\widehat{\Lambda}}(m)/7\big)_+$
$\leqslant \mathcal{C}n^{-1}\big[(1 \vee \widehat{\Lambda}_+(m_{\mathcal{C}_3}))m_{\mathcal{C}_3} + (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_5})n_{\mathcal{C}_5})\big]$;

(ii) $\sum_{m=1}^n \mathrm{pen}^{\widehat{\Lambda}}(m)\,\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m)/7\big) \leqslant \mathcal{C}n^{-1}\big[(1 \vee \widehat{\Lambda}_+(m_{\mathcal{C}_7})^2)m_{\mathcal{C}_7}^2 + (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2 n_{\mathcal{C}_8}^2)\big]$;

(iii) $\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\big) \leqslant \mathcal{C}\big[\exp\big(-\mathcal{C}_{11}\delta_{\widehat{\Lambda}}(m_-^\dagger)m_-^\dagger\big) + n^{-1}\big]$.

□

Consider now the fully data-driven aggregation of the orthogonal series estimators using either aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4) or model selection weights $\widehat{\mathbb{P}}_M^{(\infty)}$ as in eq. (3.8) combining Assumption 19 and the upper bound given in (3.23) or (3.24) we obtain the next result.

Lemma 3.2.15.
Let Assumption 19 hold true. Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\,\Delta_{\widehat{\Lambda}}(m)\,n^{-1}$, $m \in [\![1,n]\!]$, *as in Definition 38. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^n \widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\widehat{\mathbb{P}}_M^{(\eta)}$ *as in (3.4) or model selection weights* $\widehat{P}_M^{(\infty)}$ *as in eq. (3.8). Then, there is a finite numerical constant* $\mathcal{C} > 0$ *such that for all* $n, n_\lambda \in \mathbb{N}$, *for any* $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ *and associated* $m_- \in [\![1,n]\!]$ *as defined in (3.22) hold*

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^{\Lambda}(m_+^\dagger) + \tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_+^\dagger}^2(\theta^\circ) + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)$$
$$+ \mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\mathbb{P}_\varepsilon^{n_\lambda}(\mho_{m_-^\dagger}^c) + n_\lambda\mathbb{P}_\varepsilon^{n_\lambda}(\mho_{m_+^\dagger}^c)\big] + \mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) \quad (3.25)$$

□

The last bound allows us to derive an upper bound of the risk for the fully data-driven aggregated estimator in the two cases (p) and (np) introduced in section 1.6.

Theorem 3.2.3.
Let Assumption 19 hold true. Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\,\Delta_{\widehat{\Lambda}}(m)\,n^{-1}$,

$m \in [\![1,n]\!]$, as in Definition 38. Let $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ be an aggregation of the orthogonal series estimators using either aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4) or model selection weights $\mathbb{P}_M^{(\infty)}$ as in eq. (3.8).

(p) Assume there is $K \in \mathbb{N}_0$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ and $\mathfrak{b}_m(\theta^\circ) = 0$. For $K > 0$ let $c_{\theta^\circ} := \frac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}+104\kappa}{\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)} > 1$, $n_{\theta^\circ,\Lambda} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor \in \mathbb{N}$ and $n_\lambda(\theta^\circ,\Lambda) := \lfloor 289\log(K+2)\delta_\Lambda(K)\Lambda_+(K)\rfloor \in \mathbb{N}$. If $n > n_{\theta^\circ,\Lambda}$ and $n_\lambda > n_\lambda(\theta^\circ,\Lambda)$ then set $m_n^\bullet := \max\{m \in [\![K,n]\!] : n > c_{\theta^\circ}\Delta_\Lambda(m)\}$ and $m_{n_\lambda}^\bullet := \max\{m \in [\![K,n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ where the defining set, respectively, contains K and thus is not empty, and otherwise $m_n^\bullet \wedge m_{n_\lambda}^\bullet := m_{\mathcal{C}_3}\log(n\wedge n_\lambda)$. There is a numerical constant $\mathcal{C}$ and a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ given in (B.59) depending only on θ° and Λ such that for all $n, n_\lambda \in \mathbb{N}$ holds

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\big[n^{-1}\vee n_\lambda^{-1}\vee\exp\big(\tfrac{-\delta_\Lambda(m_n^\bullet\wedge m_{n_\lambda}^\bullet)m_n^\bullet\wedge m_{n_\lambda}^\bullet}{m_{\mathcal{C}_3}}\big)\big]+\mathcal{C}_{\theta^\circ,\Lambda}\{n^{-1}\vee n_\lambda^{-1}\}. \tag{3.26}$$

(np) Assume that $\mathfrak{b}_m(\theta^\circ) > 0$ for all $m \in \mathbb{N}$. Let $n_\lambda(\Lambda) := \lfloor 289\log(3)\delta_\Lambda(1)\Lambda_+(1)\rfloor \in \mathbb{N}$. If $n_\lambda > n_\lambda(\Lambda)$ then set $m_{n_\lambda}^\bullet := \max\{m \in [\![1,n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ where the defining set, respectively, contains 1 and thus is not empty. There is a numerical constant $\mathcal{C}$ such that for all $n \in \mathbb{N}$ with $m_n^\dagger := m_n^\dagger(\theta^\circ) \in [\![1,n]\!]$ as in Definition 40 and for all $n_\lambda > n_\lambda(\Lambda)$ holds

$$\begin{aligned}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) &\leqslant \mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\min_{m\in[\![1,n]\!]}\{\mathfrak{R}_n^m(\theta^\circ,\Lambda)\vee\exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_3}}\big)\}\mathbb{1}_{\{n_\lambda>n_\lambda(\Lambda)\}}\\ &\quad+\mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\{\mathfrak{b}^2_{m_n^\dagger\wedge m_{n_\lambda}^\bullet}(\theta^\circ)\vee\exp\big(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\mathcal{C}_3}}\big)\}\mathbb{1}_{\{n_\lambda>n_\lambda(\Lambda)\}}\\ &\quad+\mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)+\mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2n_\lambda^{-1}+\mathcal{C}\{\Lambda_+(m_{\mathcal{C}_3})^2m^3_{\mathcal{C}_3}+\Lambda_+(n_o)^2\}n^{-1}\end{aligned} \tag{3.27}$$

while for $n_\lambda \in [\![1,n_\lambda(\Lambda)]\!]$ we have

$$\mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)+\mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2n_\lambda^{-1}+\mathcal{C}\{\Lambda_+(m_{\mathcal{C}_3})^2m^3_{\mathcal{C}_3}+\Lambda_+(n_o)^2\}n^{-1}.$$

□

COROLLARY 3.2.3.
Let the assumptions of Theorem 3.2.3 be satisfied.

(p) If (A1) as in corollary 3.2.1 and in addition (A4) there is $n_\lambda(\theta^\circ,\Lambda) \in \mathbb{N}$ such that $\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet \geqslant m_{\mathcal{C}_3}(\log n_\lambda)$ for all $n_\lambda \geqslant n_\lambda(\theta^\circ,\Lambda)$ hold true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ depending only on θ° and Λ such that for all $n, n_\lambda \in \mathbb{N}$ holds $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}[n^{-1}\vee n_\lambda^{-1}]$.

(np) If (A2) as in corollary 3.2.1 and (A4) hold true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ depending only on θ° and Λ such that $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\{\mathfrak{R}_n^\dagger(\theta^\circ,\Lambda)+\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)+\mathfrak{b}^2_{m_{n_\lambda}^\bullet\wedge m_n^\dagger}(\theta^\circ)\}$ for all $n, n_\lambda \in \mathbb{N}$ holds true.

NUMERICAL DISCUSSION 3.2.5.
Let us briefly illustrate the last results. In case (p) the fully data-driven aggregation leads to an estimator attaining the parametric oracle rate (see Remark 1.3.1), if the additional

assumptions (**A1**) and (**A4**) are satisfied. Consider the two cases **(o)** and **(s)** for the operator Fourier sequence λ as in Num. discussion 1.3.1, where in both cases (**A1**) holds true (cf. Num. discussion 3.2.2 **(o)** and **(s)**), while

(o) $n_\lambda \sim (\log m^\bullet_{n_\lambda})\delta_\Lambda(m^\bullet_n)\Lambda_+(m^\bullet_{n_\lambda}) \sim (\log m^\bullet_{n_\lambda})(m^\bullet_{n_\lambda})^{2a}$ implies $m^\bullet_{n_\lambda} \sim (n_\lambda/\log n_\lambda)^{1/(2a)}$ and $m^\bullet_{n_\lambda}\delta_\Lambda(m^\bullet_{n_\lambda}) \sim (n_\lambda/\log n_\lambda)^{1/(2a)}$.

(s) $n_\lambda \sim (\log m^\bullet_{n_\lambda})\delta_\Lambda(m^\bullet_{n_\lambda})\Lambda_+(m^\bullet_{n_\lambda}) \sim (\log m^\bullet_{n_\lambda})(m^\bullet_{n_\lambda})^{4a}\exp((m^\bullet_{n_\lambda})^{2a})$ implies $m^\bullet_{n_\lambda} \sim (\log n_\lambda - \frac{1+4a}{2a}\log\log n_\lambda - \frac{1}{2a}\log\log\log n_\lambda)^{1/(2a)}$ and $m^\bullet_{n_\lambda}\delta_\Lambda(m^\bullet_{n_\lambda}) \sim (\log n_\lambda)^{2+1/(2a)}$.

Clearly in both cases **(o)** and **(s)** also (**A4**) is satisfied. Therefore, in this situation the fully data-driven aggregated estimator attains the parametric oracle rate. On the other hand side, in case (**np**) the fully data-driven aggregation leads to an estimator attaining the rate $\mathfrak{R}^\dagger_n(\theta^\circ,\Lambda) + \mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)$ (corollary 3.2.3), if (**A2**) and (**A4**) are satisfied and $\mathfrak{b}^2_{m^\bullet_{n_\lambda}\wedge m^\dagger_n}(\theta^\circ)$ is negligible with respect to $\mathfrak{R}^\dagger_n(\theta^\circ,\Lambda) + \mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)$, otherwise the upper bound faces a deterioration of the rate, which we illustrate considering as in Num. discussion 3.2.1 the usual behaviour **[o-o]**, **[o-s]** and **[s-o]** for the sequences $(\mathfrak{b}_m(\theta^\circ))_{m\in\mathbb{N}}$ and $(\Lambda(m))_{m\in\mathbb{N}}$. In all three cases **[o-o]**, **[o-s]** and **[s-o]** the assumption (**A4**) holds true. Moreover, in case **[o-o]**, **[o-s]** and **[s-o]** only with $p < 1/2$ the assumption (**A2**) is satisfied, and $\mathfrak{R}^\dagger_n(\theta^\circ,\Lambda)$ equals the oracle rate $\mathcal{R}^\circ_n(\theta^\circ,\Lambda)$ (cf. Num. discussion 3.2.1 **[o-o]**, **[o-s]** and **[s-o]**). In case **[o-s]** and **[s-o]** $\mathfrak{b}^2_{m^\bullet_{n_\lambda}}(\theta^\circ) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)$ while in case **[o-o]** $\mathfrak{b}^2_{m^\bullet_{n_\lambda}}(\theta^\circ) \sim (n_\lambda/\log n_\lambda)^{-p/a}$, hence

[o-o] $\mathcal{R}_{n,n_\lambda}(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\{n^{-2p/(2p+2a+1)} + n_\lambda^{-(p\wedge a)/a} + (n_\lambda/\log n_\lambda)^{-p/a}\}$

[o-s] $\mathcal{R}_{n,n_\lambda}(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\{(\log n)^{-p/a} + (\log n_\lambda)^{-p/a}\}$

[s-o] $\mathcal{R}_{n,n_\lambda}(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\{(\log n)^{(2a+1)/(2p\wedge 1)}n^{-1} + n_\lambda^{-1}\}$

Consequently, the fully data-driven estimator attains the oracle rate in case **[o-o]** with $p > a$, **[o-s]** and **[s-o]** with $p \leqslant 1/2$, while in case **[o-o]** with $p \leqslant a$ and **[s-o]** with $p > 1/2$ the rate of the fully data-driven estimator $\widehat{\theta}^{(\eta)}$ features a deterioration by a logarithmic factor $(\log n_\lambda)^{p/a}$ and $(\log n)^{(2a+1)(1-1/(2p))}$, respectively, compared to the oracle rate. □

3.2.2.2 Maximal risk bounds

By applying Lemma 3.2.11 we derive bounds for the maximal risk defined in (1.4) over ellipsoids $\Theta(\mathfrak{a},r)$ of the fully data-driven aggregated estimator $\widehat{\theta}^{(\eta)}$ using either aggregation weights $\widehat{\mathbb{P}}^{(\eta)}_M$ as in (3.4) or model selection weights $\widehat{\mathbb{P}}^{(\eta)}_M$ as in eq. (3.8). Therefore, we aim next to control the second and third right hand side term in (3.21) uniformly over $\Theta(\mathfrak{a},r)$. Results stated here are proven in appendix B.2.2.

For each $m \in \mathbb{N}$ keeping the definition 40 of $\mathfrak{R}^m_n(\mathfrak{a},\Lambda) := [\mathfrak{a}(m) \vee \Delta_\Lambda(m)\, n^{-1}]$ in mind it holds $r^2\mathfrak{R}^m_n(\mathfrak{a},\Lambda) \geqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ)$ uniformely for all $\theta^\circ \in \Theta(\mathfrak{a},r)$ and for all $m \in \mathbb{N}$. Introduce in addition $\breve{\theta}_{\overline{m}} = \sum_{s\in[\![-m,m]\!]}\lambda^+_{n_\lambda}(s)\phi(s)$. Note that $\breve{\theta}_{\overline{m}} = \Pi_{\overline{m}}\breve{\theta}_{\overline{n}}$ and $\|\Pi^\perp_{\overline{m}}\breve{\theta}_{\overline{n}}\|^2_{l^2} =$

$2\sum_{s\in\rrbracket m,n\rrbracket}\widehat{\Lambda}(s)|\phi(s)|^2$. For any $m_+^\dagger, m_-^\dagger \in \llbracket 1,n\rrbracket$ let us define

$$m_- := \min\left\{m\in\llbracket 1,m_-^\dagger\rrbracket : \|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ)\leqslant [r^2+104\kappa]\mathfrak{R}_n^{m_-^\dagger}(\mathfrak{a},\Lambda)\right\}\quad\text{and}$$
$$m_+ := \max\left\{m\in\llbracket m_+^\dagger,n\rrbracket : \mathrm{pen}^{\widehat{\Lambda}}(m)\leqslant 2[3\|\Pi^\perp_{\overline{m_+^\dagger}}\breve{\theta}_{\overline{n}}\|^2_{l^2}+2\,\mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger)]\right\}\tag{3.28}$$

where the defining set obviously contains $m_-^\dagger$ and $m_+^\dagger$, respectively, and hence, they are not empty. Keep in mind that $m_+ := m_+(\varepsilon_1,\dots,\varepsilon_{n_\lambda})$ is random but does not depend on the sample $Y_1,\dots,Y_n$.

Lemma 3.2.16.
Consider the data-driven aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4). Using the aggregation weights as in Definition 38 with $\kappa\geqslant 8\log(3e)$ and $\mho_l := \{1/4\leqslant\Lambda(s)^{-1}\widehat{\Lambda}(s)\leqslant 9/4,\ \forall\, s\in\llbracket 1,l\rrbracket\}$, $l\in\llbracket 1,n\rrbracket$, for any $m_-^\dagger,m_+^\dagger\in\llbracket 1,n\rrbracket$ and associated $m_+,m_-\in\llbracket 1,n\rrbracket$ as in (3.28) hold

(i) $\mathbb{P}_M^{(\eta)}(\llbracket 1,m_-\llbracket)\leqslant\frac{50}{\eta\kappa}\mathbb{1}_{\{m_->1\}}\exp\left(-\frac{\eta\kappa}{2}n\mathfrak{R}_n^{m_-^\dagger}(\mathfrak{a},\Lambda)\right)$
$+\,\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}}-\breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2}\geqslant\mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\}\cup\mho^c_{m_-^\dagger}}$;

(ii) $\sum_{m\in\rrbracket m_+,n\rrbracket}\mathrm{pen}^{\widehat{\Lambda}}(m)\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|^2_{l^2}<\mathrm{pen}^{\widehat{\Lambda}}(m)/7\}}\leqslant n^{-1}\{\frac{16}{\kappa\eta^2}+\frac{8}{\eta}\}$.

Keeping in mind that $\phi=\theta^\circ\cdot\lambda$. We note that uniformly for all $\theta^\circ\in\Theta(\mathfrak{a},r)$ by applying the Cauchy-Schwarz inequality holds $\|\phi\|_{l^1}\leqslant\|\lambda\|_{\mathfrak{a}}\|\theta^\circ\|_{1/\mathfrak{a}}\leqslant\|\lambda\|_{\mathfrak{a}}r$. Thereby, we obtain the next assertion immediately from Assumption 19, and we omit its elementary proof.

Lemma 3.2.17.
Assume that Assumption 19 holds true. Then, we have

(i) $\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\mathbb{E}\sum_{m=1}^{n}\mathbb{E}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|^2_{l^2}-\frac{1}{7}\mathrm{pen}^{\widehat{\Lambda}}(m)\right)_+\in\mathfrak{R}_n^\dagger(\mathfrak{a},\Lambda)$;

(ii) $\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\mathbb{E}\sum_{m=1}^{n}\mathrm{pen}^{\widehat{\Lambda}}(m)\,\mathbb{P}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|^2_{l^2}\geqslant\frac{1}{7}\mathrm{pen}^{\widehat{\Lambda}}(m)\right)\in\mathfrak{R}_n^\dagger(\mathfrak{a},\Lambda)$;

(iii) $\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)}\mathbb{E}\,\mathbb{P}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m_-^\dagger}}-\breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2}\geqslant\frac{1}{7}\mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)\right)\in\mathfrak{R}_n^\dagger(\mathfrak{a},\Lambda)$.

□

Consider now the fully data-driven aggregation of the orthogonal series estimators using either aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4) or model selection weights $\widehat{\mathbb{P}}_M^{(\infty)}$ as in eq. (3.8) combining Lemma 3.2.17 and the upper bound given in (3.23) or (3.24) we obtain the next result.

Lemma 3.2.18.
Assume that Assumption 19 holds true and consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m):=\kappa\,\Delta_{\widehat{\Lambda}}(m)\,n^{-1}$, $m\in\llbracket 1,n\rrbracket$, as in Definition 38. Let$\widehat{\theta}^{(\eta)}=\sum_{m=1}^{n}\widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ be an aggregation of the orthogonal series estimators using either aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4) or model selection weights $\mathbb{P}_M^{(\infty)}$ as in eq. (3.8). There is a finite numerical constant

$\mathcal{C} > 0$ *such that for all* $n, n_\lambda \in \mathbb{N}$*, for any* $\theta^\circ \in \Theta(\mathfrak{a}, r)$*, any* $m_-^\dagger, m_+^\dagger \in [\![1, n]\!]$ *and associated* $m_- \in [\![1, n]\!]$ *as defined in* (3.22) *hold*

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger) + \tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_+^\dagger}^2(\theta^\circ) + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_-}^2(\theta^\circ)$$
$$+ \mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\mathbb{P}_\varepsilon^{n_\lambda}(\mho_{m_-^\dagger}^c) + n_\lambda \mathbb{P}_\varepsilon^{n_\lambda}(\mho_{m_+^\dagger}^c)\big]$$
$$+ \mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ, \Lambda) + \mathcal{C}n^{-1}\{\Lambda_+(m_{\lambda,r})^2 m_{\lambda,r}^3 + \Lambda_+(n_o)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\} \quad (3.29)$$

The last bound allows us to derive an upper bound of the maximal risk over the ellipsoid $\Theta(\mathfrak{a}, r)$ for the fully data-driven aggregated estimator.

Theorem 3.2.4.
Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}$, $m \in [\![1, n]\!]$*, as in Definition 38 with numerical constant* $\kappa \geqslant 84$*. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^n \widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\widehat{\mathbb{P}}_M^{(\eta)}$ *as in* (3.4) *or model selection weights* $\mathbb{P}_M^{(\infty)}$ *as in eq.* (3.8)*. Let* $m_{\lambda,r} := \lfloor 3(400\|\lambda\|_{\mathfrak{a}} r)^2 \rfloor$ *and* $n_o := 15(600)^4$*. Let* $n_\lambda(\Lambda) := \lfloor 289 \log(3)\delta_\Lambda(1)\Lambda_+(1) \rfloor \in \mathbb{N}$*. If* $n_\lambda > n_\lambda(\Lambda)$ *then set* $m_{n_\lambda}^\bullet := \max\{m \in [\![1, n_\lambda]\!] : 289 \log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ *where the defining set, respectively, contains 1 and thus is not empty. There is a numerical constant* $\mathcal{C}$ *such that for all* $n \in \mathbb{N}$ *with* $m_n^\dagger := m_n^\dagger(\theta^\circ) \in [\![1, n]\!]$ *as in Definition 38 and for all* $n_\lambda > n_\lambda(\Lambda)$ *holds*

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}(1 \vee r^2) \min_{m \in [\![1,n]\!]} \{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda) \vee \exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\lambda,r}}\big)\}$$
$$+ \mathcal{C}(1 \vee r^2)\{\mathfrak{a}(m_n^\dagger \wedge m_{n_\lambda}^\bullet)^2 \vee \exp\big(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\lambda,r}}\big)\}$$
$$+ \mathcal{C}r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) + \mathcal{C}(1 \vee r^2)\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\lambda,r})^2 m_{\lambda,r}^3 + \Lambda_+(n_o)^2\}n^{-1} \quad (3.30)$$

while for $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$ *we have*

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda)$$
$$+ \mathcal{C}(1 \vee r^2)\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\lambda,r})^2 m_{\lambda,r}^3 + \Lambda_+(n_o)^2\}n^{-1}. \quad (3.31)$$

Corollary 3.2.4.
Let the assumptions of Theorem 3.2.4 be satisfied. If (A2) as in corollary 3.2.1 and (A4) as in corollary 3.2.3 hold true, then there is a constant $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\mathfrak{a}$*,* r *and* Λ *such that* $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\{\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) + \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) + \mathfrak{a}(m_{n_\lambda}^\bullet \wedge m_n^\dagger)^2\}$ *for all* $n, n_\lambda \in \mathbb{N}$ *holds true.*

Numerical discussion 3.2.6.
As in Num. discussion 3.2.5 shown in both cases **(o)** and **(s)** is (A4) satisfied. The fully data-driven aggregation leads to an estimator attaining the rate $\mathfrak{R}_n^\dagger(\theta^\circ, \Lambda) + \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda)$ (corollary 3.2.4), if also (A4) is satisfied and $\mathfrak{a}(m_{n_\lambda}^\bullet)^2$ is negligible with respect to $\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda)$, otherwise the upper bound faces a deterioration of the rate, which we illustrate considering as in Num. discussion 3.2.1 the usual behaviour **[o-o]**, **[o-s]** and **[s-o]** for the sequences $(\mathfrak{a}(m)^2)_{m\in\mathbb{N}}$ and $(\Lambda(m))_{m\in\mathbb{N}}$. In all three cases **[o-o]**, **[o-s]** and **[s-o]** the assumption (A4)

holds true. Moreover, in case **[o-o]**, **[o-s]** and **[s-o]** only with $p < 1/2$ the assumption **(A2)** is satisfied, and $\mathfrak{R}_n^{\dagger}(\mathfrak{a},\Lambda)$ equals the oracle rate $\mathcal{R}_n^{\star}(\mathfrak{a},\Lambda)$ (cf. Num. discussion 3.2.1 **[o-o]**, **[o-s]** and **[s-o]**). In case **[o-s]** and **[s-o]** $\mathfrak{a}(m^{\bullet}_{n_\lambda})^2 \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\mathcal{R}^{\star}_{n_\lambda}(\mathfrak{a},\Lambda)$ while in case **[o-o]** $\mathfrak{a}(m^{\bullet}_{n_\lambda})^2 \sim (n_\lambda/\log n_\lambda)^{-p/a}$, hence

[o-o] $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\Theta(\mathfrak{a},r),\Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\{n^{-2p/(2p+2a+1)} + n_\lambda^{-(p\wedge a)/a} + (n_\lambda/\log n_\lambda)^{-p/a}\}$

[o-s] $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\Theta(\mathfrak{a},r),\Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\{(\log n)^{-p/a} + (\log n_\lambda)^{-p/a}\}$

[s-o] $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\Theta(\mathfrak{a},r),\Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\{(\log n)^{(2a+1)/(2p\wedge 1)}n^{-1} + n_\lambda^{-1}\}$

Consequently, the fully data-driven estimator attains the minimax rate in case **[o-o]** with $p > a$, **[o-s]** and **[s-o]** with $p \leqslant 1/2$, while in case **[o-o]** with $p \leqslant a$ and **[s-o]** with $p > 1/2$ the rate of the fully data-driven estimator $\widehat{\theta}^{(\eta)}$ features a deterioration by a logarithmic factor $(\log n_\lambda)^{p/a}$ and $(\log n)^{(2a+1)(1-1/(2p))}$, respectively, compared to the minimax rate. □

3.3 Inverse Gaussian sequence space model

We consider the Gaussian sequence space model where, for any s in $\mathbb{N}$, we have $Y(s) \sim \mathcal{N}(\phi(s),1)$ and hence $\phi_n(s) \sim \mathcal{N}(\phi(s),n^{-1})$; and $\varepsilon(s) \sim \mathcal{N}(\lambda(s),1)$ and hence $\lambda_{n_\lambda}(s) \sim \mathcal{N}(\lambda(s),n_\lambda^{-1})$.
We will apply the strategy we just presented to this model, first with λ known, then with λ estimated. In both cases, quadratic as well as maximal risk are bounded.

3.3.1 Known operator

We assume here that we know λ and observe the vector of independent random variables Y^n. Assume now that for any s in $\mathbb{N}$, we know $\lambda(s) > 0$.

3.3.1.1 Shape of the estimator

First, let us remind that we plan to use an aggregated orthogonal series estimator $\widehat{\theta}^{(\eta)}$, with η in $\mathbb{R}_+^\star \cup \infty$ which form is reminded hereafter.

Definition We first define so-called contrast Υ and penalisation pen^Λ sequences, which allow us to define weight $\mathbb{P}_M^{(\eta)}$ on the nested sieve space (here $(\llbracket 0,m\rrbracket)_{m\in\mathbb{N}}$)

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^\Lambda : \mathbb{N} \to \mathbb{R}_-, \quad m \mapsto \mathrm{pen}^\Lambda(m);$$
$$\mathbb{P}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \frac{\exp[\eta n(\Upsilon(m)+\mathrm{pen}^\Lambda(m))]}{\sum_{k=0}^n \exp[\eta n(\Upsilon(k)+\mathrm{pen}^\Lambda(k))]}\mathbb{1}_{m\leqslant n}.$$

Notice that letting η tend to ∞ in the previous definition gives rise to the penalised contrast model selection estimator,

$$\widetilde{m} := \arg\min_{m\in\llbracket 1,n\rrbracket}\left\{\Upsilon(m) + \mathrm{pen}^\Lambda(m)\right\}$$

which corresponds to the following weights

$$\lim_{\eta\to\infty}\mathbb{P}_M^{(\eta)}(m) = \delta_{\widetilde{m}}(\{m\}) =: \mathbb{P}_M^{(\infty)}(m).$$

Here, we will use the following shape for Υ and pen^{Λ}, for $\kappa := 84$

$$\Upsilon(m) := \|\theta_{n,\overline{m}}\|_{l^2}^2; \qquad \delta_{\Lambda}(m) := \frac{\log^2(m\Lambda_+(m) \vee (m+2))}{\log^2(m+2)} \geqslant 1;$$
$$\Delta_{\Lambda}(m) := \delta_{\Lambda}(m) m \Lambda_+(m); \quad \mathrm{pen}^{\Lambda}(m) := \kappa\, \Delta_{\Lambda}(m)\, n^{-1}.$$

□

The family of estimators is hence entirely defined and can be implemented with the data we assume to have at hand in this subsection.

3.3.1.2 Oracle optimality

In a first time we are interested in the quadratic risk for any θ° fixed. Remind that the strategy we use allows to show that the following sequence is an upper bound for the quadratic risk.

Definition Remind that we defined for any θ in Θ and m in $\mathbb{N}$ the following term $\mathfrak{b}_m^2(\theta) = \|\theta_{\underline{m}}\|_{l^2} / \|\theta_{\underline{0}}\|_{l^2} \leqslant 1$. We then define a family of sequences $(\mathfrak{R}_n^m(\theta^{\circ}))_{m\in\mathbb{N}} := (\mathfrak{R}_n^m(\theta^{\circ}, \Lambda))_{m\in\mathbb{N}} = [\mathfrak{b}_m^2(\theta^{\circ}) \vee \mathrm{pen}^{\Lambda}(m)/\kappa]$ and hence it holds for all m in $[\![1, n]\!]$

$$[\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2 + \kappa]\mathfrak{R}_n^m(\theta^{\circ}) \geqslant \|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^{\circ}) \vee \mathrm{pen}^{\Lambda}(m). \tag{3.32}$$

We intend to prove that the specific choice

$$m_n^{\dagger}(\theta^{\circ}) := \arg\min\{\mathfrak{R}_n^m(\theta^{\circ}), m \in \mathbb{N}\} \in [\![1, n]\!];$$
$$\mathfrak{R}_n^{\dagger}(\theta^{\circ}) := \mathfrak{R}_n^{\dagger}(\theta^{\circ}, \Lambda) := \min\{\mathfrak{R}_n^m(\theta^{\circ}), m \in \mathbb{N}\}$$

with $\mathfrak{R}_n^{m_n^{\dagger}}(\theta^{\circ}, \Lambda) = \mathfrak{R}_n^{\dagger}(\theta^{\circ}, \Lambda)$ defines an upper bound for the convergence rate of the aggregation estimators. □

A direct application of Lemma 2.4.1 and Lemma 2.4.2 gives us the following result.

Corollary.
For any m and n in $\mathbb{N}$, we have

$$\mathbb{E}[(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|_{l^2}^2 - 12\Delta_{\Lambda}(m)n^{-1})_+] \leqslant 6\Lambda_+(m)n^{-1}\exp[-2\delta_{\Lambda}(m)m]$$
$$\mathbb{P}(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|_{l^2}^2 \geqslant 12\Delta_{\Lambda}(m)n^{-1}) \leqslant \exp[-2\delta_{\Lambda}(m)m]$$
$$\mathbb{P}(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|_{l^2}^2 \geqslant 12\mathfrak{R}_n^m(\theta^{\circ}, \Lambda)) \leqslant \exp[-2\tfrac{\mathfrak{R}_n^m(\theta^{\circ}, \Lambda)n}{\Lambda_+(m)}]$$

□

Hence, Assumption 18 is verified with constants $\mathcal{C}_1 = 1, \mathcal{C}_2 = 6, \mathcal{C}_3 = 2, \mathcal{C}_4 = 0, \mathcal{C}_6 = 1, \mathcal{C}_7 = 2, \mathcal{C}_9 = 1, \mathcal{C}_{10} = 2$, notice that $\mathcal{C}_5$, $\mathcal{C}_8$, and $\mathcal{C}_{11}$ are irrelevant here as the corresponding term is not present. We can hence apply the theorem me presented earlier.
The following theorem is then a direct application of Theorem 3.2.1 and we omit its proof.

THEOREM 3.3.1.
Consider the penalty sequence $\mathrm{pen}^{\Lambda}(m) := \kappa\,\Delta_{\Lambda}(m)\,n^{-1}$, $m \in [\![1,n]\!]$, *as in Definition 37 with numerical constant* $\kappa \geqslant 84$. *Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_{M}^{(\eta)}(m)\theta_{n,\overline{m}}$ *be an aggregation of the orthogonal series estimators, using either aggregation weights* $\mathbb{P}_{M}^{(\eta)}$ *as in (3.3), or model selection weights* $\mathbb{P}_{M}^{(\infty)}$ *as in (3.6).*

(p) Assume there is $K \in \mathbb{N}$ *with* $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^{\circ}) > 0$ *and* $\mathfrak{b}_{m}(\theta^{\circ}) = 0$. *For* $K > 0$ *let* $c_{\theta^{\circ}} := \frac{\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2 + 4\kappa}{\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^{\circ})} > 1$ *and* $n_{\theta^{\circ}} := \lfloor c_{\theta^{\circ}}\Delta_{\Lambda}(K)\rfloor \in \mathbb{N}$. *If* $n \in [\![1, n_{\theta^{\circ}}]\!]$ *then set* $m_n^{\bullet} := m_{\mathcal{C}_3}\log(n)$, *and otherwise if* $n > n_{\theta^{\circ}}$ *then set* $m_n^{\bullet} := \max\{m \in [\![K,n]\!] : n > c_{\theta^{\circ}}\Delta_{\Lambda}(m)\}$ *where the defining set contains* K *and thus is not empty. There is a finite constant* $\mathcal{C}_{\theta^{\circ},\Lambda}$ *given in (B.23) depending only on* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^{\circ}, \Lambda\big) \leqslant \mathcal{C}\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\big[n^{-1} \vee \exp\big(-2\delta_{\Lambda}(m_n^{\bullet})m_n^{\bullet}\big)\big] + \mathcal{C}_{\theta^{\circ},\Lambda}n^{-1}. \tag{3.33}$$

(np) Assume that $\mathfrak{b}_m(\theta^{\circ}) > 0$ *for all* $m \in \mathbb{N}$. *There is a finite finite constant* $\mathcal{C}_{\theta^{\circ},\Lambda}$ *given in (B.24) depending only* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^{\circ}, \Lambda\big) \leqslant \mathcal{C}(\|\theta^{\circ}_{\underline{0}}\|_{l^2}^2 \vee 1) \min_{m \in [\![1,n]\!]}\big[\mathcal{R}_n^m(\theta^{\circ}, \Lambda) \vee \exp\big(-2\delta_{\Lambda}(m)m\big)\big] + \mathcal{C}_{\theta^{\circ},\Lambda}n^{-1}. \tag{3.34}$$

□

COROLLARY 3.3.1.
Let $\kappa \geqslant 84$.

(p) If in addition (A1) there is $n_{\theta^{\circ},\Lambda} \in \mathbb{N}$ *such that* $\delta_{\Lambda}(m_n^{\bullet})m_n^{\bullet} \geqslant (\log n)/2$ *for all* $n \geqslant n_{\theta^{\circ},\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\theta^{\circ},\Lambda}$ *depending only on* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^{\circ}, \Lambda\big) \leqslant \mathcal{C}_{\theta^{\circ},\Lambda}n^{-1}$.

(np) If in addition (A2) there is $n_{\theta^{\circ},\Lambda} \in \mathbb{N}$ *such that* $m_n^{\dagger}(\theta^{\circ})\delta_{\Lambda}(m_n^{\dagger}(\theta^{\circ})) \geqslant |\log \mathfrak{R}_n^{\dagger}(\theta^{\circ}, \Lambda)|/2$ *for all* $n \geqslant n_{\theta^{\circ},\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\theta^{\circ},\Lambda}$ *depending only on* θ° *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^{\circ}, \Lambda\big) \leqslant \mathcal{C}_{\theta^{\circ},\Lambda}\mathfrak{R}_n^{\dagger}(\theta^{\circ}, \Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.*

□

In fig. 3.4 we give an illustration of the error of the aggregation estimator with $\eta = 1$, of the oracle projection estimator and of the model selection estimator, in the Gaussian sequence space model in the case of a direct problem, with the same data.
Replicating the same experiment as in fig. 3.4 many times, it allows us to estimate the distribution of the error of each estimator at a fixed value of n, which we represent in fig. 3.5.
Finally, replicating the experiment of fig. 3.5 for different values of n we can illustrate the evolution of the risk with n, as represented in fig. 3.6.

3.3.1.3 Minimax optimality

We now give interest to the maximal risk over Sobolev's ellipsoids. We aim to apply Theorem 3.2.2 which allows to show that the sequences defined hereafter are upper bounds

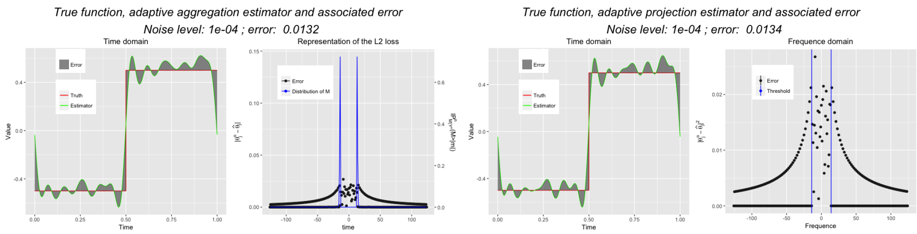

Figure 3.4: Error of the aggregation estimator and of the model selection estimator for a fixed dataset

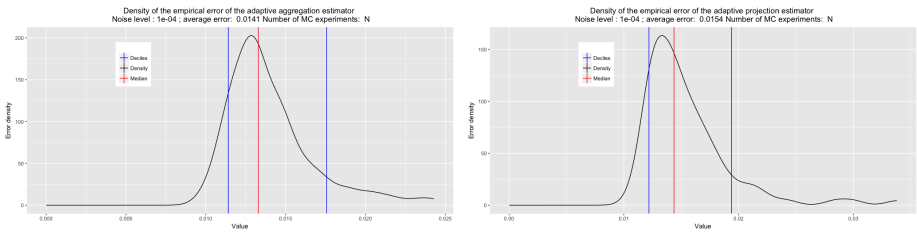

Figure 3.5: Kernel estimation of the density of the error for the aggregation estimator and the model selection estimator for a fixed true parameter θ° and noise level n

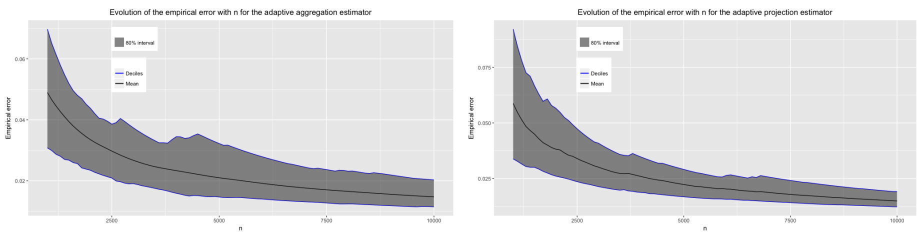

Figure 3.6: Estimation of the evolution with n of the error of the aggregation estimator and of the model selection estimator

for the maximal risk of our estimators.

DEFINITION Let be the following family of sequences, $\mathfrak{R}_n^m(\mathfrak{a}) := \mathfrak{R}_n^m(\mathfrak{a}, \Lambda) := [\mathfrak{a}(m)^2 \vee \Delta_\Lambda(m)\, n^{-1}]$. Considering the following specific case, we aim to show that it describes an upper bound for the maximal risk over $\Theta(\mathfrak{a}, r)$ for our aggregation estimator, $m_n^\dagger(\mathfrak{a}) := \arg\min\{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\} \in [\![1, n]\!]$
$\mathfrak{R}_n^\dagger(\mathfrak{a}) := \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) := \min\{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\}$ with $\mathfrak{R}_n^{m_n^\dagger(\mathfrak{a})}(\mathfrak{a}, \Lambda) = \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)$ □

The hypotheses to apply Theorem 3.2.2 are the same as for Theorem 3.2.1 and hence we directly obtain the following result.

THEOREM 3.3.2.
Assume that Assumption 18 holds true and consider the penalty sequence $\mathrm{pen}^\Lambda(m) := \kappa\, \Delta_\Lambda(m)\, n^{-1}$, $m \in [\![1, n]\!]$, *as in Definition 37. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^n \mathbb{P}_M^{(\eta)}(m)\theta_{n,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\mathbb{P}_M^{(\eta)}$ *as in* (3.3) *or model selection weights* $\mathbb{P}_M^{(\infty)}$ *as in* (3.4). *There is a finite constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *given in* (B.24) *depending only on* $\mathfrak{a}$, r *and* Λ *such that for all* $n \in \mathbb{N}$ *and for all* $m_n^\bullet \in [\![m_n^\dagger(\mathfrak{a}), n]\!]$ *with* $m_n^\dagger(\mathfrak{a}) \in [\![1, n]\!]$ *as in Definition 40 holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}(r^2 \vee 1) \min_{m \in [\![1,n]\!]} \big[\mathfrak{R}_n^m(\mathfrak{a}, \Lambda) \vee \exp\big(-2\delta_\Lambda(m) m\big)\big]\big] + \mathcal{C}_{\mathfrak{a},r,\Lambda} n^{-1}. \quad (3.35)$$

□

COROLLARY 3.3.2.
Let the assumptions of Theorem 3.2.2 be satisfied. If in addition (A) there is $n_{\mathfrak{a},r,\Lambda} \in \mathbb{N}$ *such that* $m_n^\dagger(\mathfrak{a})\delta_\Lambda(m_n^\dagger(\mathfrak{a})) \geqslant |\log \mathfrak{R}_n^\dagger(\mathfrak{a})|/2$ *for all* $n \geqslant n_{\mathfrak{a},r,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\Theta(\mathfrak{a}, r)$ *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.* □

3.3.2 Unknown operator

We now consider the case when λ is unknown and we hence use the observations $(\varepsilon_p)_{p \in [\![1, n_\lambda]\!]}$ to estimate it.

3.3.2.1 Shape of the estimator

DEFINITION We use, as usual an aggregation estimator, where, this time, the aggregating sequence does not depend on λ but on ε^{n_λ},

$$(\widehat{\theta}^{(\eta)}(s))_{s \in \mathbb{F}} = (\sum\nolimits_{m \in \mathbb{N}} \widehat{\mathbb{P}}_M^{(\eta)} \cdot \theta_{n,n_\lambda,\overline{m}}(s))_{s \in \mathbb{F}} = (\sum\nolimits_{m \geqslant |s|} \widehat{\mathbb{P}}_M^{(\eta)} \cdot \theta_{n,n_\lambda}(s))_{s \in \mathbb{F}}.$$

In particular, we give the following shape to the aggregating sequence with the contrast Υ and penalty $\mathrm{pen}^{\widehat{\Lambda}}$,

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^{\widehat{\Lambda}} : \mathbb{N} \to \mathbb{R}_-, \quad m \mapsto \mathrm{pen}^{\widehat{\Lambda}}(m);$$
$$\widehat{\mathbb{P}}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+; \quad m \mapsto \frac{\exp[\eta n(\Upsilon(m) + \mathrm{pen}^{\widehat{\Lambda}}(m))]}{\sum_{k=0}^n \exp[\eta n(\Upsilon(k) + \mathrm{pen}^{\widehat{\Lambda}}(k))]} \mathbb{1}_{m \leqslant n}.$$

Notice that if we let η tend to infinity we obtain the following penalised contrast model selection estimator,

$$\widehat{m} := \arg\min_{m \in [\![1,n]\!]} \{\Upsilon(m) + \mathrm{pen}^{\widehat{\Lambda}}(m)\}$$

which corresponds to the following weight sequence,

$$\lim_{\eta \to \infty} \widehat{\mathbb{P}}_M^{(\eta)}(m) = \delta_{\widehat{m}}(\{m\}) =: \widehat{\mathbb{P}}_M^{(\infty)}.$$

In particular, we take the following expressions for Υ and $\mathrm{pen}^{\widehat{\Lambda}}$, with $\kappa := 84$,

$$\begin{aligned}
\Upsilon(m) &:= \|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}^2; & \widehat{\Lambda}(s) &:= |\lambda_{n_\lambda}^{+}(s)|^2 \\
\widehat{\Lambda}_+(m) &:= \max\{\widehat{\Lambda}(l), l \in [\![1,m]\!]\}; & \delta_{\widehat{\Lambda}}(m) &:= \frac{\log^2(m\widehat{\Lambda}_+(m) \vee (m+2))}{\log^2(m+2)} \geqslant 1; \\
\Delta_{\widehat{\Lambda}}(m) &:= \delta_{\widehat{\Lambda}}(m) m \widehat{\Lambda}_+(m); & \mathrm{pen}^{\widehat{\Lambda}}(m) &:= \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}.
\end{aligned}$$

□

3.3.2.2 Oracle optimality

We first look at the quadratic risk for each θ° and we recall the sequence which shall be an upper bound for the quadratic risk of our estimators.

Definition Remind that we defined for any θ in Θ and m in $\mathbb{N}$ the following term $\mathfrak{b}_m^2(\theta) = \|\theta_{\underline{m}}\|_{l^2}/\|\theta_{\underline{0}}\|_{l^2} \leqslant 1$. We then define a family of sequences $(\mathfrak{R}_n^m(\theta^\circ))_{m \in \mathbb{N}} := (\mathfrak{R}_n^m(\theta^\circ, \Lambda))_{m \in \mathbb{N}} = [\mathfrak{b}_m^2(\theta^\circ) \vee \mathrm{pen}^{\Lambda}(m)/\kappa]$. We intend to prove that the specific choice

$$m_n^\dagger(\theta^\circ) := \arg\min\{\mathfrak{R}_n^m(\theta^\circ), m \in \mathbb{N}\} \in [\![1,n]\!];$$
$$\mathfrak{R}_n^\dagger(\theta^\circ) := \mathfrak{R}_n^\dagger(\theta^\circ, \Lambda) := \min\{\mathfrak{R}_n^m(\theta^\circ), m \in \mathbb{N}\}$$

with $\mathfrak{R}_n^{m_n^\dagger}(\theta^\circ, \Lambda) = \mathfrak{R}_n^\dagger(\theta^\circ, \Lambda)$ defines an upper bound for the convergence rate of the aggregation estimators. □

Our method gives us a simple assumption to verify in order to obtain this result.
The following result, which is a direct application, once again, of Lemma 2.4.1 and Lemma 2.4.2 gives us precisely what we want.

Corollary.
For any m, n, and n_λ in $\mathbb{N}$ we have,

$$\begin{aligned}
\mathbb{E}_{\phi_n|\lambda_{n_\lambda}}[(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 - 12\Delta_{\widehat{\Lambda}}(m)n^{-1})_+] &\leqslant 6n^{-1}\widehat{\Lambda}_+(m)\exp[-2\delta_{\widehat{\Lambda}}(m)]; \\
\mathbb{P}_{\phi_n|\lambda_{n_\lambda}}[\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}] &\leqslant \exp[-2\delta_{\widehat{\Lambda}}(m)]; \\
\mathbb{P}(|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3) &\leqslant \exp[-\tfrac{n_\lambda}{6\Lambda(s)}].
\end{aligned}$$

□

The following theorem is then a direct consequence of Theorem 3.2.3 and we omit its proof.

Theorem 3.3.3.
Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}$, $m \in [\![1,n]\!]$, *as in Definition 38. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\widehat{\mathbb{P}}_M^{(\eta)}$ *as in (3.4) or model selection weights* $\mathbb{P}_M^{(\infty)}$ *as in eq. (3.8).*

(p) Assume there is $K \in \mathbb{N}_0$ *with* $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ *and* $\mathfrak{b}_m(\theta^\circ) = 0$. *For* $K > 0$ *let* $c_{\theta^\circ} := \frac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}+104\kappa}{\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)} > 1$, $n_{\theta^\circ,\Lambda} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor \in \mathbb{N}$ *and* $n_\lambda(\theta^\circ,\Lambda) := \lfloor 289\log(K+2)\delta_\Lambda(K)\Lambda_+(K)\rfloor \in \mathbb{N}$. *If* $n > n_{\theta^\circ,\Lambda}$ *and* $n_\lambda > n_\lambda(\theta^\circ,\Lambda)$ *then set* $m_n^\bullet := \max\{m \in [\![K,n]\!] : n > c_{\theta^\circ}\Delta_\Lambda(m)\}$ *and* $m_{n_\lambda}^\bullet := \max\{m \in [\![K,n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ *where the defining set, respectively, contains* K *and thus is not empty, and otherwise* $m_n^\bullet \wedge m_{n_\lambda}^\bullet := m_{\mathcal{C}_3}\log(n\wedge n_\lambda)$. *There is a numerical constant* $\mathcal{C}$ *and a constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *given in (B.59) depending only on* θ° *and* Λ *such that for all* $n, n_\lambda \in \mathbb{N}$ *holds*

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\big[n^{-1}\vee n_\lambda^{-1}\vee\exp\big(\tfrac{-\delta_\Lambda(m_n^\bullet\wedge m_{n_\lambda}^\bullet)m_n^\bullet\wedge m_{n_\lambda}^\bullet}{m_{\mathcal{C}_3}}\big)\big]+\mathcal{C}_{\theta^\circ,\Lambda}\{n^{-1}\vee n_\lambda^{-1}\}. \tag{3.36}$$

(np) Assume that $\mathfrak{b}_m(\theta^\circ) > 0$ *for all* $m \in \mathbb{N}$. *Let* $n_\lambda(\Lambda) := \lfloor 289\log(3)\delta_\Lambda(1)\Lambda_+(1)\rfloor \in \mathbb{N}$. *If* $n_\lambda > n_\lambda(\Lambda)$ *then set* $m_{n_\lambda}^\bullet := \max\{m \in [\![1,n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ *where the defining set, respectively, contains 1 and thus is not empty. There is a numerical constant* $\mathcal{C}$ *such that for all* $n \in \mathbb{N}$ *with* $m_n^\dagger := m_n^\dagger(\theta^\circ) \in [\![1,n]\!]$ *as in Definition 40 and for all* $n_\lambda > n_\lambda(\Lambda)$ *holds*

$$\begin{aligned}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) &\leqslant \mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\min_{m\in[\![1,n]\!]}\{\mathfrak{R}_n^m(\theta^\circ,\Lambda)\vee\exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_3}}\big)\}\mathbb{1}_{\{n_\lambda>n_\lambda(\Lambda)\}}\\ &\quad+\mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\{\mathfrak{b}^2_{m_n^\dagger\wedge m_{n_\lambda}^\bullet}(\theta^\circ)\vee\exp\big(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\mathcal{C}_3}}\big)\}\mathbb{1}_{\{n_\lambda>n_\lambda(\Lambda)\}}\\ &\quad+\mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)+\mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2n_\lambda^{-1}+\mathcal{C}\{\Lambda_+(m_{\mathcal{C}_3})^2m^3_{\mathcal{C}_3}+\Lambda_+(n_o)^2\}n^{-1}\end{aligned} \tag{3.37}$$

while for $n_\lambda \in [\![1,n_\lambda(\Lambda)]\!]$ *we have*

$\mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)+\mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2n_\lambda^{-1}+\mathcal{C}\{\Lambda_+(m_{\mathcal{C}_3})^2m^3_{\mathcal{C}_3}+\Lambda_+(n_o)^2\}n^{-1}.$

□

Corollary 3.3.3. *(p) If (A1) as in corollary 3.2.1 and in addition (A4) there is* $n_\lambda(\theta^\circ,\Lambda) \in \mathbb{N}$ *such that* $\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet \geqslant m_{\mathcal{C}_3}(\log n_\lambda)$ *for all* $n_\lambda \geqslant n_\lambda(\theta^\circ,\Lambda)$ *hold true, then there is a constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *depending only on* θ° *and* Λ *such that for all* $n, n_\lambda \in \mathbb{N}$ *holds* $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}[n^{-1}\vee n_\lambda^{-1}]$.

(np) If (A2) as in corollary 3.2.1 and (A4) hold true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ *depending only on* θ° *and* Λ *such that* $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\{\mathfrak{R}_n^\dagger(\theta^\circ,\Lambda)+\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda)+\mathfrak{b}^2_{m_{n_\lambda}^\bullet\wedge m_n^\dagger}(\theta^\circ)\}$ *for all* $n, n_\lambda \in \mathbb{N}$ *holds true.*

3.3.2.3 Minimax optimality

We now give interest to the maximal risk over Sobolev's ellipsoids. We aim to apply Theorem 3.2.2 which allows to show that the sequences defined hereafter are upper bounds for the maximal risk of our estimators.

DEFINITION Let be the following family of sequences, $\mathfrak{R}_n^m(\mathfrak{a}) := \mathfrak{R}_n^m(\mathfrak{a}, \Lambda) := [\mathfrak{a}(m)^2 \vee \Delta_\Lambda(m)\, n^{-1}]$. Considering the following specific case, we aim to show that it describes an upper bound for the maximal risk over $\Theta(\mathfrak{a}, r)$ for our aggregation estimator, $m_n^\dagger(\mathfrak{a}) := \arg\min\{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\} \in [\![1, n]\!]$
$\mathfrak{R}_n^\dagger(\mathfrak{a}) := \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) := \min\{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda), m \in \mathbb{N}\}$ with $\mathfrak{R}_n^{m_n^\dagger(\mathfrak{a})}(\mathfrak{a}, \Lambda) = \mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda)$ □

The hypotheses to apply Theorem 3.2.2 are the same as for Theorem 3.2.1 and hence we directly obtain the following result.

THEOREM 3.3.4.
Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}$, $m \in [\![1, n]\!]$, *as in Definition 38 with numerical constant* $\kappa \geqslant 84$. *Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\widehat{\mathbb{P}}_M^{(\eta)}$ *as in (3.4) or model selection weights* $\mathbb{P}_M^{(\infty)}$ *as in eq. (3.8). Let* $m_{\lambda,r} := \lfloor 3(400\|\lambda\|_{\mathfrak{a}} r)^2 \rfloor$ *and* $n_o := 15(600)^4$. *Let* $n_\lambda(\Lambda) := \lfloor 289 \log(3)\delta_\Lambda(1)\Lambda_+(1) \rfloor \in \mathbb{N}$. *If* $n_\lambda > n_\lambda(\Lambda)$ *then set* $m_{n_\lambda}^\bullet := \max\{m \in [\![1, n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ *where the defining set, respectively, contains 1 and thus is not empty. There is a numerical constant* $\mathcal{C}$ *such that for all* $n \in \mathbb{N}$ *with* $m_n^\dagger := m_n^\dagger(\theta^\circ) \in [\![1, n]\!]$ *as in Definition 38 and for all* $n_\lambda > n_\lambda(\Lambda)$ *holds*

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}(1 \vee r^2) \min_{m \in [\![1,n]\!]} \{\mathfrak{R}_n^m(\mathfrak{a}, \Lambda) \vee \exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\lambda,r}}\big)\}$$
$$+ \mathcal{C}(1 \vee r^2)\{\mathfrak{a}(m_n^\dagger \wedge m_{n_\lambda}^\bullet)^2 \vee \exp\big(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\lambda,r}}\big)\}$$
$$+ \mathcal{C}r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) + \mathcal{C}(1 \vee r^2)\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\lambda,r})^2 m_{\lambda,r}^3 + \Lambda_+(n_o)^2\}n^{-1} \quad (3.38)$$

while for $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$ *we have*

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}r^2\mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) + \mathcal{C}(1 \vee r^2)\Lambda_+(1)^2 n_\lambda^{-1}$$
$$+ \mathcal{C}\{\Lambda_+(m_{\lambda,r})^2 m_{\lambda,r}^3 + \Lambda_+(n_o)^2\}n^{-1}. \quad (3.39)$$

COROLLARY 3.3.4.
Let the assumptions of Theorem 3.2.4 be satisfied. If (A2) as in corollary 3.2.1 and (A4) as in corollary 3.2.3 hold true, then there is a constant $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\mathfrak{a}$, r *and* Λ *such that* $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\{\mathfrak{R}_n^\dagger(\mathfrak{a}, \Lambda) + \mathcal{R}_{n_\lambda}^\star(\mathfrak{a}, \Lambda) + \mathfrak{a}(m_{n_\lambda}^\bullet \wedge m_n^\dagger)^2\}$ *for all* $n, n_\lambda \in \mathbb{N}$ *holds true.*

3.4 Circular deconvolution model

We place ourselves in the framework introduced in section 1.6. We will first consider the case of a known operator, first with independent data, then with an absolutely regular process. We will then consider an unknown operator. In all cases, we start by giving the explicit, detailed expression of the estimator, as well as the sequence which shall be proven o be an upper bound for the quadratic risk, before reminding the hypotheses to prove in order to obtain the convergence, we shall then proceed to show that the said hypothesis holds true before concluding. The proofs for the results obtained in this section may be found in appendix C.

3.4.1 Independent data and known convolution density

We place ourselves under Assumption 3 and Assumption 5 and intend to apply the strategy highlighted in section 3.2.1.

We assume here that we know λ and observe the vector of independent random variables Y^n. Assume now that for any s in $\mathbb{N}$, we know $\lambda(s)$ and $\lambda(s) > 0$.

3.4.1.1 Shape of the estimator

First, let us remind that we plan to use an aggregated orthogonal series estimator $\widehat{\theta}^{(\eta)}$, with η in $\mathbb{R}_+^\star \cup \infty$ which form is reminded hereafter.

Definition We first define so-called contrast Υ and penalisation pen^Λ sequences, which allow us to define weight $\mathbb{P}_M^{(\eta)}$ on the nested sieve space (here $(\llbracket 0, m \rrbracket)_{m \in \mathbb{N}}$)

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^\Lambda : \mathbb{N} \to \mathbb{R}_-, \quad m \mapsto \mathrm{pen}^\Lambda(m);$$
$$\mathbb{P}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \frac{\exp[\eta n(\Upsilon(m)+\mathrm{pen}^\Lambda(m))]}{\sum_{k=0}^{n} \exp[\eta n(\Upsilon(k)+\mathrm{pen}^\Lambda(k))]} \mathbb{1}_{m \leqslant n}.$$

Notice that letting η tend to ∞ in the previous definition gives rise to the penalised contrast model selection estimator,

$$\widetilde{m} := \arg\min_{m \in \llbracket 1, n \rrbracket} \left\{ \Upsilon(m) + \mathrm{pen}^\Lambda(m) \right\}$$

which corresponds to the following weights

$$\lim_{\eta \to \infty} \mathbb{P}_M^{(\eta)}(m) = \delta_{\widetilde{m}}(\{m\}) =: \mathbb{P}_M^{(\infty)}(m).$$

Here, we will use the following shape for Υ and pen^Λ, for $\kappa := 84$

$$\begin{aligned} \Upsilon(m) &:= \|\theta_{n,\overline{m}}\|_{l^2}^2; & \delta_\Lambda(m) &:= \frac{\log^2(m\Lambda_+(m) \vee (m+2))}{\log^2(m+2)} \geqslant 1; \\ \Delta_\Lambda(m) &:= \delta_\Lambda(m) m \Lambda_+(m); & \mathrm{pen}^\Lambda(m) &:= \kappa\, \Delta_\Lambda(m)\, n^{-1}. \end{aligned}$$

□

The family of estimators is hence entirely defined and can be implemented with the data we assume to have at hand in this subsection.

3.4.1.2 Quadratic risk bounds of the aggregated estimator

In a first time we are interested in the quadratic risk for any θ° fixed. Remind that the strategy we use allows to show that the following sequence is an upper bound for the quadratic risk.

Definition Remind that we defined for any θ in Θ and m in $\mathbb{N}$ the following term $\mathfrak{b}_m^2(\theta) = \|\theta_{\underline{m}}\|_{l^2}/\|\theta_{\underline{0}}\|_{l^2} \leqslant 1$. We then define a family of sequences $(\mathfrak{R}_n^m(\theta^\circ))_{m\in\mathbb{N}} := (\mathfrak{R}_n^m(\theta^\circ,\Lambda))_{m\in\mathbb{N}} = [\mathfrak{b}_m^2(\theta^\circ) \vee \mathrm{pen}^\Lambda(m)/\kappa]$ and hence it holds for all m in $[\![1,n]\!]$

$$[\|\theta^\circ_{\underline{0}}\|_{l^2}^2 + \kappa]\mathfrak{R}_n^m(\theta^\circ) \geqslant \|\theta^\circ_{\underline{0}}\|_{l^2}^2\, \mathfrak{b}_m^2(\theta^\circ) \vee \mathrm{pen}^\Lambda(m). \tag{3.40}$$

We intend to prove that the specific choice

$$m_n^\dagger(\theta^\circ) := \arg\min\{\mathfrak{R}_n^m(\theta^\circ), m\in\mathbb{N}\} \in [\![1,n]\!];$$
$$\mathfrak{R}_n^\dagger(\theta^\circ) := \mathfrak{R}_n^\dagger(\theta^\circ,\Lambda) := \min\{\mathfrak{R}_n^m(\theta^\circ), m\in\mathbb{N}\}$$

with $\mathfrak{R}_n^{m_n^\dagger}(\theta^\circ,\Lambda) = \mathfrak{R}_n^\dagger(\theta^\circ,\Lambda)$ defines an upper bound for the convergence rate of the aggregation estimators. □

The following lemma shows that the assumption for our method is verified.

Lemma 3.4.1.
Let $\Lambda_\circ(m) = \frac{1}{m}\sum_{s\in[\![1,m]\!]}\Lambda(s)$, $\Lambda_+(m) = \max\{\Lambda(s), s\in[\![1,m]\!]\}$, $\delta_\Lambda(m) \geqslant 1$ and $\Delta_\Lambda(m) = \delta_\Lambda(m) m\Lambda_+(m)$, then there is a numerical constant $\mathcal{C}$ such that for all $n\in\mathbb{N}$ and $m\in[\![1,n]\!]$ holds

(i) $\mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|_{l^2}^2 - 12\Delta_\Lambda(m)n^{-1}\right)_+$
$$\leqslant \mathcal{C}\left[\frac{\|\phi\|_{l^1}\Lambda_+(m)}{n}\exp\left(\frac{-\delta_\Lambda(m)m}{3\|\phi\|_{l^1}}\right) + \frac{2m\Lambda_+(m)}{n^2}\exp\left(\frac{-\sqrt{n\delta_\Lambda(m)}}{200}\right)\right]$$

(ii) $\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|_{l^2}^2 \geqslant 12\Delta_\Lambda(m)n^{-1}\right) \leqslant 3\left[\exp\left(\frac{-\delta_\Lambda(m)m}{200\|\phi\|_{l^1}}\right) + \exp\left(\frac{-\sqrt{n\delta_\Lambda(m)}}{200}\right)\right]$

(iii) $\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|_{l^2}^2 \geqslant 12\mathfrak{R}_n^m(\theta^\circ,\Lambda)\right) \leqslant 3\left[\exp\left(\frac{-n\mathfrak{R}_n^m(\theta^\circ,\Lambda)}{200\|\phi\|_{l^1}\Lambda_+(m)}\right) + \exp\left(\frac{-n\sqrt{\mathfrak{R}_n^m(\theta^\circ,\Lambda)}}{200\sqrt{m\Lambda_+(m)}}\right)\right]$

Hence, using Theorem 3.2.1 gives us the following result.

Theorem 3.4.1.
Consider the penalty sequence $\mathrm{pen}^\Lambda(m) := \kappa\,\Delta_\Lambda(m)\,n^{-1}$, $m\in[\![1,n]\!]$, as in Definition 37 with numerical constant $\kappa \geqslant 84$. Let $\widehat{\theta}^{(\eta)} = \sum_{m=1}^n \mathbb{P}_M^{(\eta)}(m)\theta_{n,\overline{m}}$ be an aggregation of the orthogonal series estimators, using either aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3), or model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.6).

(p) Assume there is $K\in\mathbb{N}$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ and $\mathfrak{b}_m(\theta^\circ) = 0$. For $K > 0$ let $c_{\theta^\circ} := \frac{\|\theta^\circ_{\underline{0}}\|_{l^2}^2 + 4\kappa}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ)} > 1$ and $n_{\theta^\circ} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor \in \mathbb{N}$. If $n\in[\![1,n_{\theta^\circ}]\!]$ then set $m_n^\bullet := m_{\mathcal{C}_3}\log(n)$, and otherwise if $n > n_{\theta^\circ}$ then set $m_n^\bullet := \max\{m\in[\![K,n]\!] : n > c_{\theta^\circ}\Delta_\Lambda(m)\}$ where the defining set contains K and thus is not empty. There is a finite constant $\mathcal{C}_{\theta^\circ,\Lambda}$ given in (B.23) depending only on θ° and Λ such that for all $n\in\mathbb{N}$ holds

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\big[n^{-1} \vee \exp\big(-2\delta_\Lambda(m_n^\bullet)m_n^\bullet\big)\big] + \mathcal{C}_{\theta^\circ,\Lambda}n^{-1}. \tag{3.41}$$

(np) Assume that $\mathfrak{b}_m(\theta^\circ) > 0$ for all $m\in\mathbb{N}$. There is a finite finite constant $\mathcal{C}_{\theta^\circ,\Lambda}$ given

in (B.24) *depending only* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}(\|\theta^\circ_{\underline{0}}\|^2_{l^2} \vee 1) \min_{m \in [\![1,n]\!]} \big[\mathcal{R}^m_n(\theta^\circ, \Lambda) \vee \exp\big(-2\delta_\Lambda(m)m\big)\big] + \mathcal{C}_{\theta^\circ,\Lambda} n^{-1}. \tag{3.42}$$

□

Corollary 3.4.1.
Let $\kappa \geqslant 84$.

(p) If in addition (A1) there is $n_{\theta^\circ,\Lambda} \in \mathbb{N}$ *such that* $\delta_\Lambda(m^\bullet_n)m^\bullet_n \geqslant (\log n)/2$ *for all* $n \geqslant n_{\theta^\circ,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *depending only on* θ° *and* Λ *such that for all* $n \in \mathbb{N}$ *holds* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda} n^{-1}$.

(np) If in addition (A2) there is $n_{\theta^\circ,\Lambda} \in \mathbb{N}$ *such that* $m^\dagger_n(\theta^\circ)\delta_\Lambda(m^\dagger_n(\theta^\circ)) \geqslant |\log \mathfrak{R}^\dagger_n(\theta^\circ, \Lambda)|/2$ *for all* $n \geqslant n_{\theta^\circ,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\theta^\circ,\Lambda}$ *depending only on* θ° *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\mathfrak{R}^\dagger_n(\theta^\circ, \Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.*

□

3.4.1.3 Maximal risk bounds of the aggregated estimator

We now give interest to the maximal risk over Sobolev's ellipsoids. We aim to apply Theorem 3.2.2 which allows to show that the sequences defined hereafter are upper bounds for the maximal risk of our estimators.

Definition Let be the following family of sequences, $\mathfrak{R}^m_n(\mathfrak{a}) := \mathfrak{R}^m_n(\mathfrak{a}, \Lambda) := [\mathfrak{a}(m)^2 \vee \Delta_\Lambda(m)\, n^{-1}]$. Considering the following specific case, we aim to show that it describes an upper bound for the maximal risk over $\Theta(\mathfrak{a}, r)$ for our aggregation estimator, $m^\dagger_n(\mathfrak{a}) := \arg\min\{\mathfrak{R}^m_n(\mathfrak{a}, \Lambda), m \in \mathbb{N}\} \in [\![1, n]\!]$
$\mathfrak{R}^\dagger_n(\mathfrak{a}) := \mathfrak{R}^\dagger_n(\mathfrak{a}, \Lambda) := \min\{\mathfrak{R}^m_n(\mathfrak{a}, \Lambda), m \in \mathbb{N}\}$ with $\mathfrak{R}^{m^\dagger_n(\mathfrak{a})}_n(\mathfrak{a}, \Lambda) = \mathfrak{R}^\dagger_n(\mathfrak{a}, \Lambda)$ □

The hypotheses to apply Theorem 3.2.2 are the same as for Theorem 3.2.1 and hence we directly obtain the following result.

Theorem 3.4.2.
Consider the penalty sequence $\mathrm{pen}^\Lambda(m) := \kappa\, \Delta_\Lambda(m)\, n^{-1}$, $m \in [\![1, n]\!]$, *as in Definition 37. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^n \mathbb{P}^{(\eta)}_M(m)\theta_{n,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\mathbb{P}^{(\eta)}_M$ *as in* (3.3) *or model selection weights* $\mathbb{P}^{(\infty)}_M$ *as in* (3.4). *There is a finite constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *given in* (B.24) *depending only on* $\mathfrak{a}$, r *and* Λ *such that for all* $n \in \mathbb{N}$ *and for all* $m^\bullet_n \in [\![m^\dagger_n(\mathfrak{a}), n]\!]$ *with* $m^\dagger_n(\mathfrak{a}) \in [\![1, n]\!]$ *as in Definition 40 holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}(r^2 \vee 1) \min_{m \in [\![1,n]\!]} \big[\mathfrak{R}^m_n(\mathfrak{a}, \Lambda) \vee \exp\big(-2\delta_\Lambda(m)m\big)]\big] + \mathcal{C}_{\mathfrak{a},r,\Lambda} n^{-1}. \tag{3.43}$$

□

Corollary 3.4.2.
Let the assumptions of Theorem 3.2.2 be satisfied. If in addition (A) there is $n_{\mathfrak{a},r,\Lambda} \in \mathbb{N}$

such that $m_n^\dagger(\mathfrak{a})\delta_\Lambda(m_n^\dagger(\mathfrak{a})) \geqslant |\log \mathfrak{R}_n^\dagger(\mathfrak{a})|/2$ *for all* $n \geqslant n_{\mathfrak{a},r,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\Theta(\mathfrak{a},r)$ *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)},\Theta(\mathfrak{a},r),\Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\mathfrak{R}_n^\dagger(\mathfrak{a},\Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.* □

3.4.2 Absolutely regular process and and known noise density

We now replace the independence assumption Assumption 5 by the absolute regularity assumption Assumption 6, that is, we recall

Assumption Considering the process of observations $(Y_p)_{p\in\mathbb{Z}}$, assume that for any p, the joint distribution $\mathbb{P}_{Y_0,Y_p}$ of Y_0 and Y_p admits a density denoted g_{Y_0,Y_p} which is square integrable. Denote $\|x\|^2_{L^{2,2}} := \int\int_{[0,1]^2} |x(t,t')|^2 \,\mathrm{d}t\,\mathrm{d}t'$ the L^2-norm for functions of two variables and for any t and t' in $[0,1]$ set $(x\otimes x)(t,t') = x(t)\cdot x(t')$. Then, we assume $\gamma_g := \sup_{p\geqslant 1}\|g_{Y_0^n,Y_p^n} - g\otimes g\|_{L^{2,2}} < \infty$. □

Under this assumption we will use Lemma 1.2.3, which is

Lemma.
Let the process of observations $(Y_p)_{p\in\mathbb{N}}$ *be a strictly stationary process with associated sequence of mixing coefficients* $(\beta(Y_0,Y_p))_{p\in\mathbb{N}}$. *Under Assumption 6, for any* $n \geqslant 1$*;* m *and* l *in* $\mathbb{N}$ *with* $m \leqslant l$ *and* $K \in [\![0, n-1]\!]$*, it holds*

$$\sum_{m\leqslant|s|\leqslant l}\mathbb{V}[\sum_{p=1}^{n} e_s(Y_p)] \\ \leqslant n2(l-m+1)\{1+2[\gamma_g K(l-m+1)^{-1/2} + 2\sum_{p=K+1}^{n-1}\beta(Y_0,Y_p)]\}.$$

Moreover, as $\sum_{p\in\mathbb{N}}\beta(Y_0,Y_p)$ *is finite, we have* $\lim_{K arrow \infty}\sum_{p=K+1}^{\infty}\beta(Y_0,Y_p) = 0$*, so we can find* K° *in* $\mathbb{N}$ *such that for any* K *greater than* K°*,* $\sum_{p=K+1}^{\infty}\beta(Y_0,Y_p) \leqslant \frac{1}{4}$*. We can take* $K = \frac{\sqrt{l-m+1}}{4\gamma_g}$ *and assuming that this choice is greater than* K°*, we have*

$$\sum_{m\leqslant|s|\leqslant l}\mathbb{V}[\sum_{p=1}^{n} e_s(Y_p)] \leqslant 4n(l-m+1).$$

□

And we will also use Lemma 1.2.4, recalled bellow.

Lemma.
Assume that the universe is rich enough in the sense that there exist a sequence of random variables with uniform distribution on $[0,1]$ *which is independent of* $(Y_p)_{p\in\mathbb{Z}}$.
Then, there exist a sequence $(Y_p^\perp)_{p\in\mathbb{Z}}$ *satisfying the following properties. For any positive integer* w *and for any strictly positive integer* q*, define the sets* $(I^e_{q,p})_{p\in[\![1,w]\!]} := [\![2(q-1)w+1, (2q-1)w]\!]$ *and* $(I^o_{q,p})_{p\in[\![1,w]\!]} := [\![(2q-1)w+1, 2qw]\!]$.
Define for any q *in* $\mathbb{Z}$ *the vectors of random variables* $E_q := (Y^n_{I^e_{q,p}})_{p\in[\![1,w]\!]}$*;* $O_q := (Y^n_{I^o_{q,p}})_{p\in[\![1,w]\!]}$*; and their counterparts* $E_q^\perp := (Y^{n,\perp}_{I^e_{q,p}})_{p\in[\![1,w]\!]}$ *and* $O_q^\perp := (Y^{n,\perp}_{I^o_{q,p}})_{p\in[\![1,w]\!]}$.
Then, $(Y_p^\perp)_{p\in\mathbb{Z}}$ *satisfies:*

- *for any integer* q*,* $E_q^\perp$*,* E_q*,* $O_q^\perp$*, and* O_q *are identically distributed;*

- *for any integer* q, $\mathbb{P}(E_q \neq E_q^{\perp}) \leqslant \beta_w$ *and* $\mathbb{P}(O_q \neq O_q^{\perp}) \leqslant \beta_w$;
- $(E_q^{\perp})_{q\in\mathbb{Z}}$ *are independent and identically distributed and* $(O_q^{\perp})_{q\in\mathbb{Z}}$ *as well.*

□

3.4.2.1 Shape of the estimator

Note that we will use the same estimator as previously and hence no knowledge about $(\beta_p)_{w\in\mathbb{Z}}$ is required. We recall here the shape of the estimator.

Definition We first define so-called contrast Υ and penalisation pen^{Λ} sequences, which allow us to define weight $\mathbb{P}_M^{(\eta)}$ on the nested sieve space (here $(\llbracket 0,m\rrbracket)_{m\in\mathbb{N}}$)

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^{\Lambda} : \mathbb{N} \to \mathbb{R}_-, \quad m \mapsto \mathrm{pen}^{\Lambda}(m);$$
$$\mathbb{P}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \frac{\exp[\eta n(\Upsilon(m)+\mathrm{pen}^{\Lambda}(m))]}{\sum_{k=0}^{n}\exp[\eta n(\Upsilon(k)+\mathrm{pen}^{\Lambda}(k))]}\mathbb{1}_{m\leqslant n}.$$

Notice that letting η tend to ∞ in the previous definition gives rise to the penalised contrast model selection estimator,

$$\widetilde{m} := \arg\min_{m\in\llbracket 1,n\rrbracket}\left\{\Upsilon(m) + \mathrm{pen}^{\Lambda}(m)\right\}$$

which corresponds to the following weights

$$\lim_{\eta\to\infty}\mathbb{P}_M^{(\eta)}(m) = \delta_{\widetilde{m}}(\{m\}) =: \mathbb{P}_M^{(\infty)}(m).$$

Here, we will use the following shape for Υ and pen^{Λ}, for $\kappa := 84$

$$\begin{aligned}\Upsilon(m) &:= \|\theta_{n,\overline{m}}\|_{l^2}^2; & \delta_{\Lambda}(m) &:= \frac{\log^2(m\Lambda_+(m)\vee(m+2))}{\log^2(m+2)} \geqslant 1;\\ \Delta_{\Lambda}(m) &:= \delta_{\Lambda}(m)m\Lambda_+(m); & \mathrm{pen}^{\Lambda}(m) &:= \kappa\,\Delta_{\Lambda}(m)\,n^{-1}.\end{aligned}$$

□

3.4.2.2 Oracle optimality

We will use the same technic as previously. Let us hence recall the rate which we use.

Definition Remind that we defined for any θ in Θ and m in $\mathbb{N}$ the following term $\mathfrak{b}_m^2(\theta) = \|\theta_{\underline{m}}\|_{l^2}/\|\theta_{\underline{0}}\|_{l^2} \leqslant 1$. We then define a family of sequences $(\mathfrak{R}_n^m(\theta^\circ))_{m\in\mathbb{N}} := (\mathfrak{R}_n^m(\theta^\circ,\Lambda))_{m\in\mathbb{N}} = [\mathfrak{b}_m^2(\theta^\circ) \vee \mathrm{pen}^{\Lambda}(m)/\kappa]$. We intend to prove that the specific choice

$$m_n^{\dagger}(\theta^\circ) := \arg\min\{\mathfrak{R}_n^m(\theta^\circ), m\in\mathbb{N}\} \in \llbracket 1,n\rrbracket;$$
$$\mathfrak{R}_n^{\dagger}(\theta^\circ) := \mathfrak{R}_n^{\dagger}(\theta^\circ,\Lambda) := \min\{\mathfrak{R}_n^m(\theta^\circ), m\in\mathbb{N}\}$$

with $\mathfrak{R}_n^{m_n^{\dagger}}(\theta^\circ,\Lambda) = \mathfrak{R}_n^{\dagger}(\theta^\circ,\Lambda)$ defines an upper bound for the convergence rate of the aggregation estimators. □

Verifying our hypotheses is technically more demanding than in the independent case, we hence give some more details in this section.

As previously, we will apply Talagrand's inequalities presented in Lemma C.0.1. However, due to the dependence structure of our data, they cannot be applied directly and we first use Lemma 1.2.4. It allows us to obtain the following result.

Lemma 3.4.2.
For any integer k and l such that $k \leqslant l$, define for any t in $\mathbb{B}_{k,l}$ the functional $\overline{\nu}_t = \langle t|\theta_n - \theta^\circ\rangle_{l^2}$. Under Lemma 1.2.4, we define

$$\begin{aligned}\overline{\nu}_t^{e,\perp} &= r^{-1}\sum\nolimits_{q=1}^{r}(v_t(E_q^\perp) - \mathbb{E}[v_t(E_q^\perp)]); \quad v_t(E_q^\perp) = s^{-1}\sum\nolimits_{p=1}^{s}\nu_t(E_{q,p}^\perp);\\ \nu_t(E_{q,p}^\perp) &= \sum\nolimits_{k\leqslant|s|\leqslant l}(t(s)\overline{\lambda(s)}^{-1}e_s(E_{q,p}^\perp)).\end{aligned}$$

Then, for any sequence $(C_n)_{n\in\mathbb{N}}$, we have the following inequalities:

$$\begin{aligned}\mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\langle t|\theta_n - \theta^\circ\rangle_{l^2}|^2 - C_n)_+] \leqslant& 2\cdot\mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\overline{\nu}_t^{e,\perp}|^2 - C_n)_+]\\ &+ 2\cdot\mathbb{E}[\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\overline{\nu}_t^{e,\perp} - \overline{\nu}_t^{e}|^2] \qquad (3.44)\\ \mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\langle t|\theta_n - \theta^\circ\rangle_{l^2}| \geqslant C_n) \leqslant& 2\,\mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\overline{\nu}_t^{e,\perp}| \geqslant C_n)\\ &+ 2\sum_{q=1}^{r}\mathbb{P}(\{E_q^\perp \neq E_q\}) \qquad (3.45)\end{aligned}$$

□

We now apply Lemma C.0.1 in the respective first parts of eq. (3.44) and eq. (3.45).

Lemma 3.4.3.
For any integers k and l with $k < l$; consider a triplet h^2, H^2 and v verifying

$$\begin{aligned}h^2 &\geqslant \sum\nolimits_{k\leqslant|s|\leqslant l}\Lambda(s); \quad H^2 \geqslant n^{-1}\Lambda_+(l)(l-k+1)(\delta_\Lambda(m)+1);\\ v &\geqslant 4w^{-1}\sqrt{m}\Lambda_+(m)\sqrt{2\|\phi\|_{l^1}\sum\nolimits_{p=1}^{\infty}(p+1)\beta_p};\end{aligned}$$

then, under Assumption 6, for any $C > 0$, we have:

$$\begin{aligned}\mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\overline{\nu}_t^{e,\perp}|^2 - 6H^2)_+] &\leqslant C[\tfrac{v}{r}\exp(\tfrac{-rH^2}{6v}) + \tfrac{h^2}{r^2}\exp(\tfrac{-rH}{100h})]; \qquad (3.46)\\ \mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{k,l}}|\overline{\nu}_t^{e,\perp}| \geqslant 6H^2) &\leqslant 3(\exp[-\tfrac{rH^2}{400v}] + \exp[\tfrac{-rH}{100h}]). \qquad (3.47)\end{aligned}$$

□

Considering the results we obtained in Lemma 3.4.3 and Lemma 3.4.2 we obtain by combining 3.44 and 3.46, using the fact that $\delta_\Lambda(m) \geqslant 1$, we select

$$l = 1; \quad k = m; \quad h^2 = m\Lambda_+(m) \geqslant m\Lambda_\circ(m);$$
$$v = 4w^{-1}\sqrt{m}\Lambda_+(m)\sqrt{2\|\phi\|_{l^1}\sum\nolimits_{p=1}^{\infty}(p+1)\beta_p};$$
$$H^2 = 2\Delta_\Lambda(m)n^{-1} = n^{-1}m\Lambda_+(m)(2\delta_\Lambda(m)) \geqslant n^{-1}\Lambda_+(m)m(\delta_\Lambda(m)+1);$$

then, given the constant $\mathcal{C}_{\beta,\phi} := \sqrt{2\|\phi\|_{l^1}\sum_{p=1}^{\infty}(p+1)\beta_p}$ we have

$$\mathbb{E}[(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} - 12n^{-1}\Lambda_+(m)m\delta_\Lambda(m))_+] \leqslant C[\mathcal{C}_{\beta,\phi}\tfrac{8\sqrt{m}\Lambda_+(m)}{n}\exp(\tfrac{-\sqrt{m}\delta_\Lambda(m)}{24\mathcal{C}_{\beta,\phi}}) + \tfrac{m\Lambda_+(m)}{r^2}\exp(\tfrac{-\sqrt{2n\delta_\Lambda(m)}}{200w})];$$

and by combining 3.45 and eq. (3.47)

$$\mathbb{P}(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12n^{-1}\Lambda_+(m)m\delta_\Lambda(m)) \leqslant 3(\exp[-\tfrac{\sqrt{m}\delta_\Lambda(m)}{1600\mathcal{C}_{\beta,\phi}}] + \exp[\tfrac{-\sqrt{n\delta_\Lambda(m)}}{100w}];$$

and if we chose

$$l = 1; \quad k = m; \quad h^2 = m\Lambda_+(m) \geqslant m\Lambda_\circ(m); \quad v = 4w^{-1}\sqrt{m}\Lambda_+(m)\sqrt{2\|\phi\|_{l^1}\sum\nolimits_{p=1}^{\infty}(p+1)\beta_p};$$
$$H^2 = 2\mathfrak{R}^m_n(\theta^\circ, \Lambda) = 2[\mathfrak{b}^2_m(\theta^\circ) \vee \delta_\Lambda(m)m\Lambda_+(m)n^{-1}] \geqslant n^{-1}\Lambda_+(m)m(\delta_\Lambda(m)+1);$$

we obtain

$$\mathbb{P}(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\mathfrak{R}^m_n(\theta^\circ, \Lambda)) \leqslant 3(\exp[-\tfrac{n\mathfrak{R}^m_n(\theta^\circ,\Lambda)}{1600\sqrt{m}\Lambda_+(m)\mathcal{C}_{\beta,\phi}}] + \exp[\tfrac{-n\sqrt{2\mathfrak{R}^m_n(\theta^\circ,\Lambda)}}{200w\sqrt{m\Lambda_+(m)}}])$$

From this we deduce the following lemma.

Lemma 3.4.4.
For any n in $\mathbb{N}$ we have

$$\sum_{m=1}^{n} \mathbb{E}[(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|^2_{l^2} - \mathrm{pen}^{\Lambda}/7)_+]$$
$$\leqslant Cn^{-1}[\mathcal{C}_{\beta,\phi} 8 \sum_{m=1}^{n} \sqrt{m}\Lambda_+(m)\exp(\frac{-\sqrt{m}\delta_\Lambda(m)}{24\mathcal{C}_{\beta,\phi}}) + \frac{4q}{r}\sum_{m=1}^{n} m\Lambda_+(m)\exp(\frac{-\sqrt{2n\delta_\Lambda(m)}}{200w})] \quad (3.48)$$

$$\sum_{m=1}^{n} \frac{\mathrm{pen}^{\Lambda}(m)}{7}\mathbb{P}(\|\theta_{n,\overline{m}} - \theta^{\circ}_{\overline{m}}\|^2_{l^2} \geqslant \mathrm{pen}^{\Lambda}(m)/7)$$
$$\leqslant 3(\sum_{m=1}^{n} \frac{\mathrm{pen}^{\Lambda}(m)}{7}\exp[-\frac{\sqrt{m}\delta_\Lambda(m)}{1600\mathcal{C}_{\beta,\phi}}] + \sum_{m=1}^{n} \frac{\mathrm{pen}^{\Lambda}(m)}{7}\exp[\frac{-\sqrt{n\delta_\Lambda(m)}}{100w}] \quad (3.49)$$

$$\mathbb{P}(\|\theta_{n,\overline{m^{\dagger}_{-}}} - \theta^{\circ}_{\overline{m^{\dagger}_{-}}}\|^2_{l^2} \geqslant 12\mathfrak{R}^m_n(\theta^{\circ},\Lambda))$$
$$\leqslant 3(\exp[-\frac{n\mathfrak{R}^m_n(\theta^{\circ},\Lambda)}{1600\sqrt{m}\Lambda_+(m)\mathcal{C}_{\beta,\phi}}] + \exp[\frac{-n\sqrt{2\mathfrak{R}^m_n(\theta^{\circ},\Lambda)}}{200w\sqrt{m\Lambda_+(m)}}] \quad (3.50)$$

□

We finally control the respective second parts of 3.44 and 3.45 using the properties of section 1.2.4.

Lemma 3.4.5.
For any integers k and l with $k \leqslant l$

$$\mathbb{E}[\sup_{t\in\mathbb{B}_{k,l}} |\overline{\nu}^{e,\perp}_t - \overline{\nu}^{e}_t|^2] \leqslant 2r\beta_w \sum_{k\leqslant|s|\leqslant l} \Lambda(s); \quad (3.51)$$
$$\sum_{q=1}^{r} \mathbb{P}(\{E^{\perp}_q \neq E_q\}) \leqslant r\beta_w. \quad (3.52)$$

□

We see that, in order to avoid a degradation of the rate, we need to make a stronger assumption on the sequence β_w. A sufficient condition is suggested hereafter.

Assumption 20 Assume that the sequence of mixing coefficients $(\beta_w)_{w\in\mathbb{N}}$ is such that there exists a numerical constant $\mathcal{C}$ such that for all n in $\mathbb{N}$ and m in $[\![1,n]\!]$ we have

$$n^2\,\mathrm{pen}^{\Lambda}(m)\beta_w \leqslant \mathcal{C}.$$

□

Under this assumption, the main claim follows.

Theorem 3.4.3.
Assume that Assumption 20 holds true and consider the penalty sequence $\mathrm{pen}^{\Lambda}(m) :=$

$\kappa\,\Delta_\Lambda(m)\,n^{-1}$, $m \in [\![1,n]\!]$, as in Definition 37 with numerical constant $\kappa \geqslant 84$. Let $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_M^{(\eta)}(m)\theta_{n,\overline{m}}$ be an aggregation of the orthogonal series estimators, using either aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3), or model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.6).

(p) Assume there is $K \in \mathbb{N}$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ and $\mathfrak{b}_m(\theta^\circ) = 0$. For $K > 0$ let $c_{\theta^\circ} := \frac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}+4\kappa}{\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)} > 1$ and $n_{\theta^\circ} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor \in \mathbb{N}$. If $n \in [\![1, n_{\theta^\circ}]\!]$ then set $m_n^\bullet := m_{\mathcal{C}_3}\log(n)$, and otherwise if $n > n_{\theta^\circ}$ then set $m_n^\bullet := \max\{m \in [\![K,n]\!] : n > c_{\theta^\circ}\Delta_\Lambda(m)\}$ where the defining set contains K and thus is not empty. There is a finite constant $\mathcal{C}_{\theta^\circ,\Lambda}$ given in (B.23) depending only on θ° and Λ such that for all $n \in \mathbb{N}$ holds

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\big[n^{-1} \vee \exp\big(-2\delta_\Lambda(m_n^\bullet)m_n^\bullet\big)\big] + \mathcal{C}_{\theta^\circ,\Lambda}n^{-1}. \tag{3.53}$$

(np) Assume that $\mathfrak{b}_m(\theta^\circ) > 0$ for all $m \in \mathbb{N}$. There is a finite finite constant $\mathcal{C}_{\theta^\circ,\Lambda}$ given in (B.24) depending only θ° and Λ such that for all $n \in \mathbb{N}$ holds

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}(\|\theta^\circ_{\underline{0}}\|^2_{l^2} \vee 1) \min_{m\in[\![1,n]\!]} \big[\mathcal{R}_n^m(\theta^\circ, \Lambda) \vee \exp\big(-2\delta_\Lambda(m)m\big)\big] + \mathcal{C}_{\theta^\circ,\Lambda}n^{-1}. \tag{3.54}$$

□

Corollary 3.4.3.
Let $\kappa \geqslant 84$.

(p) If in addition (A1) there is $n_{\theta^\circ,\Lambda} \in \mathbb{N}$ such that $\delta_\Lambda(m_n^\bullet)m_n^\bullet \geqslant (\log n)/2$ for all $n \geqslant n_{\theta^\circ,\Lambda}$ holds true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ depending only on θ° and Λ such that for all $n \in \mathbb{N}$ holds $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}n^{-1}$.

(np) If in addition (A2) there is $n_{\theta^\circ,\Lambda} \in \mathbb{N}$ such that $m_n^\dagger(\theta^\circ)\delta_\Lambda(m_n^\dagger(\theta^\circ)) \geqslant |\log \mathfrak{R}_n^\dagger(\theta^\circ,\Lambda)|/2$ for all $n \geqslant n_{\theta^\circ,\Lambda}$ holds true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ depending only on θ° and Λ such that $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \theta^\circ, \Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\mathfrak{R}_n^\dagger(\theta^\circ,\Lambda)$ for all $n \in \mathbb{N}$ holds true.

□

Note that Assumption 20 adds a strong constraint on the dependence compared to the basic definition of the absolutely regular processes. It would be of interest to investigate the convergence rate of our estimator when this hypothesis is relaxed. It is however beyong the scope of this thesis.

3.4.2.3 Maximal risk bounds of the aggregated estimator

We now give interest to the maximal risk over Sobolev's ellipsoids. We aim to apply Theorem 3.2.2 which allows to show that the sequences defined hereafter are upper bounds for the maximal risk of our estimators.

Definition Let be the following family of sequences, $\mathfrak{R}_n^m(\mathfrak{a}) := \mathfrak{R}_n^m(\mathfrak{a},\Lambda) := [\mathfrak{a}(m)^2 \vee \Delta_\Lambda(m)\,n^{-1}]$. Considering the following specific case, we aim to show that it describes an upper bound for the maximal risk over $\Theta(\mathfrak{a}, r)$ for our aggregation estimator, $m_n^\dagger(\mathfrak{a}) := \arg\min\{\mathfrak{R}_n^m(\mathfrak{a},\Lambda), m \in \mathbb{N}\} \in [\![1,n]\!]$
$\mathfrak{R}_n^\dagger(\mathfrak{a}) := \mathfrak{R}_n^\dagger(\mathfrak{a},\Lambda) := \min\{\mathfrak{R}_n^m(\mathfrak{a},\Lambda), m \in \mathbb{N}\}$ with $\mathfrak{R}_n^{m_n^\dagger(\mathfrak{a})}(\mathfrak{a},\Lambda) = \mathfrak{R}_n^\dagger(\mathfrak{a},\Lambda)$ □

The hypotheses to apply Theorem 3.2.2 are the same as for Theorem 3.2.1 and hence we directly obtain the following result.

THEOREM 3.4.4.
Assume that Assumption 20 holds true and consider the penalty sequence $\mathrm{pen}^{\Lambda}(m) := \kappa\,\Delta_{\Lambda}(m)\,n^{-1}$, $m \in [\![1,n]\!]$, *as in Definition 37. Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \mathbb{P}_M^{(\eta)}(m)\theta_{n,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\mathbb{P}_M^{(\eta)}$ *as in* (3.3) *or model selection weights* $\mathbb{P}_M^{(\infty)}$ *as in* (3.4). *There is a finite constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *given in* (B.24) *depending only on* $\mathfrak{a}$, r *and* Λ *such that for all* $n \in \mathbb{N}$ *and for all* $m_n^{\bullet} \in [\![m_n^{\dagger}(\mathfrak{a}), n]\!]$ *with* $m_n^{\dagger}(\mathfrak{a}) \in [\![1,n]\!]$ *as in Definition 40 holds*

$$\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a},r), \Lambda\big) \leqslant \mathcal{C}(r^2 \vee 1) \min_{m\in[\![1,n]\!]} \big[\mathfrak{R}_n^{m}(\mathfrak{a},\Lambda) \vee \exp\left(-2\delta_{\Lambda}(m)m\right)])\big] + \mathcal{C}_{\mathfrak{a},r,\Lambda} n^{-1}. \quad (3.55)$$

□

COROLLARY 3.4.4.
Let the assumptions of Theorem 3.2.2 be satisfied. If in addition **(A)** *there is* $n_{\mathfrak{a},r,\Lambda} \in \mathbb{N}$ *such that* $m_n^{\dagger}(\mathfrak{a})\delta_{\Lambda}(m_n^{\dagger}(\mathfrak{a})) \geqslant |\log \mathfrak{R}_n^{\dagger}(\mathfrak{a})|/2$ *for all* $n \geqslant n_{\mathfrak{a},r,\Lambda}$ *holds true, then there is a constant* $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\Theta(\mathfrak{a},r)$ *and* Λ *such that* $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a},r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\mathfrak{R}_n^{\dagger}(\mathfrak{a},\Lambda)$ *for all* $n \in \mathbb{N}$ *holds true.* □

Note that once again Assumption 20 is required to obtain the result.

3.4.3 Independent data and unknown noise density

We now consider the case when λ is unknown and we hence use the observations $(\varepsilon_p)_{p\in[\![1,n_\lambda]\!]}$ to estimate it.

3.4.3.1 Shape of the estimator

DEFINITION We use, as usual an aggregation estimator, where, this time, the aggregating sequence does not depend on λ but on ε^{n_λ},

$$(\widehat{\theta}^{(\eta)}(s))_{s\in\mathbb{F}} = (\sum\nolimits_{m\in\mathbb{N}} \widehat{\mathbb{P}}_M^{(\eta)} \cdot \theta_{n,n_\lambda,\overline{m}}(s))_{s\in\mathbb{F}} = (\sum\nolimits_{m\geqslant|s|} \widehat{\mathbb{P}}_M^{(\eta)} \cdot \theta_{n,n_\lambda}(s))_{s\in\mathbb{F}}.$$

In particular, we give the following shape to the aggregating sequence with the contrast Υ and penalty $\mathrm{pen}^{\widehat{\Lambda}}$,

$$\Upsilon : \mathbb{N} \to \mathbb{R}_+, \quad m \mapsto \Upsilon(m); \qquad \mathrm{pen}^{\widehat{\Lambda}} : \mathbb{N} \to \mathbb{R}_-, \quad m \mapsto \mathrm{pen}^{\widehat{\Lambda}}(m);$$
$$\widehat{\mathbb{P}}_M^{(\eta)} : \mathbb{N} \to \mathbb{R}_+; \quad m \mapsto \frac{\exp[\eta n(\Upsilon(m)+\mathrm{pen}^{\widehat{\Lambda}}(m))]}{\sum_{k=0}^{n}\exp[\eta n(\Upsilon(k)+\mathrm{pen}^{\widehat{\Lambda}}(k))]}\mathbb{1}_{m\leqslant n}.$$

Notice that if we let η tend to infinity we obtain the following penalised contrast model selection estimator,

$$\widehat{m} := \arg\min_{m\in[\![1,n]\!]} \left\{\Upsilon(m) + \mathrm{pen}^{\widehat{\Lambda}}(m)\right\}$$

which corresponds to the following weight sequence,

$$\lim_{\eta\to\infty} \widehat{\mathbb{P}}^{(\eta)}_M(m) = \delta_{\widehat{m}}(\{m\}) =: \widehat{\mathbb{P}}^{(\infty)}_M.$$

In particular, we take the following expressions for Υ and $\mathrm{pen}^{\widehat{\Lambda}}$, with $\kappa := 84$,

$$\begin{aligned} \Upsilon(m) &:= \|\theta_{n,n_\lambda,\overline{m}}\|^2_{l^2}; & \widehat{\Lambda}(s) &:= |\lambda^+_{n_\lambda}(s)|^2 \\ \widehat{\Lambda}_+(m) &:= \max\{\widehat{\Lambda}(l), l \in [\![1,m]\!]\}; & \delta_{\widehat{\Lambda}}(m) &:= \tfrac{\log^2(m\widehat{\Lambda}_+(m)\vee(m+2))}{\log^2(m+2)} \geqslant 1; \\ \Delta_{\widehat{\Lambda}}(m) &:= \delta_{\widehat{\Lambda}}(m) m \widehat{\Lambda}_+(m); & \mathrm{pen}^{\widehat{\Lambda}}(m) &:= \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}. \end{aligned}$$

□

3.4.3.2 Risk bounds of the aggregated estimator

We first look at the quadratic risk for each θ° and we recall the sequence which shall be an upper bound for the quadratic risk of our estimators.

Definition Remind that we defined for any θ in Θ and m in $\mathbb{N}$ the following term $\mathfrak{b}^2_m(\theta) = \|\theta_{\underline{m}}\|_{l^2}/\|\theta_{\underline{0}}\|_{l^2} \leqslant 1$. We then define a family of sequences $(\mathfrak{R}^m_n(\theta^\circ))_{m\in\mathbb{N}} := (\mathfrak{R}^m_n(\theta^\circ,\Lambda))_{m\in\mathbb{N}} = [\mathfrak{b}^2_m(\theta^\circ) \vee \mathrm{pen}^{\Lambda}(m)/\kappa]$. We intend to prove that the specific choice

$$m^\dagger_n(\theta^\circ) := \arg\min\{\mathfrak{R}^m_n(\theta^\circ), m \in \mathbb{N}\} \in [\![1,n]\!];$$
$$\mathfrak{R}^\dagger_n(\theta^\circ) := \mathfrak{R}^\dagger_n(\theta^\circ,\Lambda) := \min\{\mathfrak{R}^m_n(\theta^\circ), m \in \mathbb{N}\}$$

with $\mathfrak{R}^{m^\dagger_n}_n(\theta^\circ,\Lambda) = \mathfrak{R}^\dagger_n(\theta^\circ,\Lambda)$ defines an upper bound for the convergence rate of the aggregation estimators. □

The following result assures that the hypotheses to apply our method are verified.

Lemma 3.4.6.
Consider $\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}} = \sum_{|s|\in[\![1,m]\!]} \lambda^+_{n_\lambda}(s)(\phi_n(s) - \phi(s))e_s$. Conditionally on $\{\varepsilon_1,\dots,\varepsilon_{n_\lambda}\}$ the r.v.'s $\{Y_1,\dots,Y_n\}$ are iid. and we denote by $\mathbb{P}_{Y|\varepsilon}$ and $\mathbb{E}_{Y|\varepsilon}$ their conditional distribution and expectation, respectively. Let $\widehat{\Lambda}(s) = |\lambda^+_{n_\lambda}(s)|^2$, $\Lambda_\circ(m) = \frac{1}{m}\sum_{s\in[\![1,m]\!]}\widehat{\Lambda}(s)$, $\widehat{\Lambda}_+(m) = \max\{\widehat{\Lambda}(s), s \in [\![1,m]\!]\}$, $\Delta_{\widehat{\Lambda}}(m) = \delta_{\widehat{\Lambda}}(m) m \widehat{\Lambda}_+(m)$ and $\delta_{\widehat{\Lambda}}(m) \geqslant 1$. Then there is a numerical constant $\mathcal{C}$ such that for all $n \in \mathbb{N}$ and $m \in [\![1,n]\!]$ holds

(i) $$\mathbb{E}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{l^2} - 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\right)_+ \leqslant \mathcal{C}\left[\frac{\|\phi\|_{l^1}\widehat{\Lambda}_+(m)}{n}\exp\left(\frac{-\delta_{\widehat{\Lambda}}(m)m}{3\|\phi\|_{l^1}}\right) + \frac{2m\widehat{\Lambda}_+(m)}{n^2}\exp\left(\frac{-\sqrt{n\delta_{\widehat{\Lambda}}(m)}}{200}\right)\right]$$

(ii) $$\mathbb{P}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{l^2} \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\right) \leqslant 3\left[\exp\left(\frac{-\delta_{\widehat{\Lambda}}(m)m}{200\|\phi\|_{l^1}}\right) + \exp\left(\frac{-\sqrt{n\delta_{\widehat{\Lambda}}(m)}}{200}\right)\right]$$

(iii) $$\mathbb{P}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{l^2} \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\right) \leqslant 3\left[\exp\left(\frac{-\delta_{\widehat{\Lambda}}(m)m}{200\|\phi\|_{l^1}}\right) + \exp\left(\frac{-n\sqrt{\mathfrak{R}^m_n(\theta^\circ,\widehat{\Lambda})}}{200\sqrt{m\widehat{\Lambda}_+(m)}}\right)\right]$$

□

Lemma 3.4.7.
Given $m \in \mathbb{N}$ for all $s \in [\![1, m]\!]$ we have

$$\mathbb{P}\big(|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\big) \leqslant 2\exp\big(-\tfrac{n_\lambda|\lambda(s)|^2}{72}\big) \leqslant 2\exp\big(-\tfrac{n_\lambda}{72\Lambda_+(m)}\big). \tag{3.56}$$

The following theorem is then a direct consequence of Theorem 3.2.3 and we omit its proof.

Theorem 3.4.5.
Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\,\Delta_{\widehat{\Lambda}}(m)\,n^{-1}$, $m \in [\![1, n]\!]$, as in Definition 38. Let $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \widehat{\mathbb{P}}_M^{(\eta)}(m)\theta_{n,n_\lambda,\overline{m}}$ be an aggregation of the orthogonal series estimators using either aggregation weights $\widehat{\mathbb{P}}_M^{(\eta)}$ as in (3.4) or model selection weights $\mathbb{P}_M^{(\infty)}$ as in eq. (3.8).

(p) Assume there is $K \in \mathbb{N}_0$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ and $\mathfrak{b}_m(\theta^\circ) = 0$. For $K > 0$ let $c_{\theta^\circ} := \frac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}+104\kappa}{\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)} > 1$, $n_{\theta^\circ,\Lambda} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor \in \mathbb{N}$ and $n_\lambda(\theta^\circ,\Lambda) := \lfloor 289\log(K + 2)\delta_\Lambda(K)\Lambda_+(K)\rfloor \in \mathbb{N}$. If $n > n_{\theta^\circ,\Lambda}$ and $n_\lambda > n_\lambda(\theta^\circ,\Lambda)$ then set $m_n^\bullet := \max\{m \in [\![K, n]\!] : n > c_{\theta^\circ}\Delta_\Lambda(m)\}$ and $m_{n_\lambda}^\bullet := \max\{m \in [\![K, n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ where the defining set, respectively, contains K and thus is not empty, and otherwise $m_n^\bullet \wedge m_{n_\lambda}^\bullet := m_{\mathcal{C}_3}\log(n \wedge n_\lambda)$. There is a numerical constant $\mathcal{C}$ and a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ given in (B.59) depending only on θ° and Λ such that for all $n, n_\lambda \in \mathbb{N}$ holds

$$\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\big[n^{-1} \vee n_\lambda^{-1} \vee \exp\big(\tfrac{-\delta_\Lambda(m_n^\bullet \wedge m_{n_\lambda}^\bullet)m_n^\bullet \wedge m_{n_\lambda}^\bullet}{m_{\mathcal{C}_3}}\big)\big] + \mathcal{C}_{\theta^\circ,\Lambda}\{n^{-1} \vee n_\lambda^{-1}\}. \tag{3.57}$$

(np) Assume that $\mathfrak{b}_m(\theta^\circ) > 0$ for all $m \in \mathbb{N}$. Let $n_\lambda(\Lambda) := \lfloor 289\log(3)\delta_\Lambda(1)\Lambda_+(1)\rfloor \in \mathbb{N}$. If $n_\lambda > n_\lambda(\Lambda)$ then set $m_{n_\lambda}^\bullet := \max\{m \in [\![1, n_\lambda]\!] : 289\log(m + 2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ where the defining set, respectively, contains 1 and thus is not empty. There is a numerical constant $\mathcal{C}$ such that for all $n \in \mathbb{N}$ with $m_n^\dagger := m_n^\dagger(\theta^\circ) \in [\![1, n]\!]$ as in Definition 40 and for all $n_\lambda > n_\lambda(\Lambda)$ holds

$$\begin{aligned}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) &\leqslant \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2}) \min_{m\in[\![1,n]\!]}\{\mathfrak{R}_n^m(\theta^\circ,\Lambda) \vee \exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_3}}\big)\}\mathbb{1}_{\{n_\lambda > n_\lambda(\Lambda)\}}\\ &\quad + \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\{\mathfrak{b}^2_{m_n^\dagger \wedge m_{n_\lambda}^\bullet}(\theta^\circ) \vee \exp\big(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\mathcal{C}_3}}\big)\}\mathbb{1}_{\{n_\lambda > n_\lambda(\Lambda)\}}\\ &+ \mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\mathcal{C}_3})^2 m^3_{\mathcal{C}_3} + \Lambda_+(n_o)^2\}n^{-1}\end{aligned} \tag{3.58}$$

while for $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$ we have

$\mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\mathcal{C}_3})^2 m^3_{\mathcal{C}_3} + \Lambda_+(n_o)^2\}n^{-1}$.

□

Corollary 3.4.5. *(p) If (A1) as in corollary 3.2.1 and in addition (A4) there is $n_\lambda(\theta^\circ,\Lambda) \in \mathbb{N}$ such that $\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet \geqslant m_{\mathcal{C}_3}(\log n_\lambda)$ for all $n_\lambda \geqslant n_\lambda(\theta^\circ,\Lambda)$ hold true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ depending only on θ° and Λ such that for all $n, n_\lambda \in \mathbb{N}$ holds $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}[n^{-1} \vee n_\lambda^{-1}]$.*

(np) If (A2) as in corollary 3.2.1 and (A4) hold true, then there is a constant $\mathcal{C}_{\theta^\circ,\Lambda}$ depending only on θ° and Λ such that $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)},\theta^\circ,\Lambda\big) \leqslant \mathcal{C}_{\theta^\circ,\Lambda}\{\mathfrak{R}_n^\dagger(\theta^\circ,\Lambda) + \mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) +$

$\mathfrak{b}^2_{m^\bullet_{n_\lambda}\wedge m^\dagger_n}(\theta^\circ)\}$ *for all* $n, n_\lambda \in \mathbb{N}$ *holds true.*

We hence see that we obtained the optimal rate under mild assumptions for the circular density deconvolution model.

3.4.3.3 Maximal risk bounds of the aggregated estimator

We now give interest to the maximal risk over Sobolev's ellipsoids. We aim to apply Theorem 3.2.2 which allows to show that the sequences defined hereafter are upper bounds for the maximal risk of our estimators.

Definition Let be the following family of sequences, $\mathfrak{R}^m_n(\mathfrak{a}) := \mathfrak{R}^m_n(\mathfrak{a}, \Lambda) := [\mathfrak{a}(m)^2 \vee \Delta_\Lambda(m)\, n^{-1}]$. Considering the following specific case, we aim to show that it describes an upper bound for the maximal risk over $\Theta(\mathfrak{a}, r)$ for our aggregation estimator, $m^\dagger_n(\mathfrak{a}) := \arg\min\{\mathfrak{R}^m_n(\mathfrak{a}, \Lambda), m \in \mathbb{N}\} \in [\![1, n]\!]$
$\mathfrak{R}^\dagger_n(\mathfrak{a}) := \mathfrak{R}^\dagger_n(\mathfrak{a}, \Lambda) := \min\{\mathfrak{R}^m_n(\mathfrak{a}, \Lambda), m \in \mathbb{N}\}$ with $\mathfrak{R}^{m^\dagger_n(\mathfrak{a})}_n(\mathfrak{a}, \Lambda) = \mathfrak{R}^\dagger_n(\mathfrak{a}, \Lambda)$ □

The hypotheses to apply Theorem 3.2.4 are the same as for Theorem 3.2.1 and hence we directly obtain the following result.

Theorem 3.4.6.
Consider the penalty sequence $\mathrm{pen}^{\widehat{\Lambda}}(m) := \kappa\, \Delta_{\widehat{\Lambda}}(m)\, n^{-1}$, $m \in [\![1, n]\!]$, *as in Definition 38 with numerical constant* $\kappa \geqslant 84$. *Let* $\widehat{\theta}^{(\eta)} = \sum_{m=1}^{n} \widehat{\mathbb{P}}^{(\eta)}_M(m)\theta_{n,n_\lambda,\overline{m}}$ *be an aggregation of the orthogonal series estimators using either aggregation weights* $\widehat{\mathbb{P}}^{(\eta)}_M$ *as in (3.4) or model selection weights* $\mathbb{P}^{(\infty)}_M$ *as in eq. (3.8). Let* $m_{\lambda,r} := \lfloor 3(400\|\lambda\|_{\mathfrak{a}} r)^2 \rfloor$ *and* $n_o := 15(600)^4$. *Let* $n_\lambda(\Lambda) := \lfloor 289\log(3)\delta_\Lambda(1)\Lambda_+(1)\rfloor \in \mathbb{N}$. *If* $n_\lambda > n_\lambda(\Lambda)$ *then set* $m^\bullet_{n_\lambda} := \max\{m \in [\![1, n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ *where the defining set, respectively, contains 1 and thus is not empty. There is a numerical constant* $\mathcal{C}$ *such that for all* $n \in \mathbb{N}$ *with* $m^\dagger_n := m^\dagger_n(\theta^\circ) \in [\![1, n]\!]$ *as in Definition 38 and for all* $n_\lambda > n_\lambda(\Lambda)$ *holds*

$$\begin{aligned}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}(1 \vee r^2) \min_{m \in [\![1,n]\!]} \{\mathfrak{R}^m_n(\mathfrak{a}, \Lambda) \vee \exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\lambda,r}}\big)\}\\ + \mathcal{C}(1 \vee r^2)\{\mathfrak{a}(m^\dagger_n \wedge m^\bullet_{n_\lambda})^2 \vee \exp\big(\tfrac{-\delta_\Lambda(m^\bullet_{n_\lambda})m^\bullet_{n_\lambda}}{m_{\lambda,r}}\big)\}\\ + \mathcal{C}r^2\mathcal{R}^\star_{n_\lambda}(\mathfrak{a}, \Lambda) + \mathcal{C}(1 \vee r^2)\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\lambda,r})^2 m^3_{\lambda,r} + \Lambda_+(n_o)^2\}n^{-1}\end{aligned} \quad (3.59)$$

while for $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$ *we have*

$$\begin{aligned}\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}r^2\mathcal{R}^\star_{n_\lambda}(\mathfrak{a}, \Lambda) + \mathcal{C}(1 \vee r^2)\Lambda_+(1)^2 n_\lambda^{-1}\\ + \mathcal{C}\{\Lambda_+(m_{\lambda,r})^2 m^3_{\lambda,r} + \Lambda_+(n_o)^2\}n^{-1}.\end{aligned} \quad (3.60)$$

Corollary 3.4.6.
Let the assumptions of Theorem 3.2.4 be satisfied. If (A2) as in corollary 3.2.1 and (A4) as in corollary 3.2.3 hold true, then there is a constant $\mathcal{C}_{\mathfrak{a},r,\Lambda}$ *depending only on* $\mathfrak{a}$, r *and* Λ *such that* $\mathcal{R}_{n,n_\lambda}\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\{\mathfrak{R}^\dagger_n(\mathfrak{a}, \Lambda) + \mathcal{R}^\star_{n_\lambda}(\mathfrak{a}, \Lambda) + \mathfrak{a}(m^\bullet_{n_\lambda} \wedge m^\dagger_n)^2\}$ *for all* $n, n_\lambda \in \mathbb{N}$ *holds true.*

Consequently, the fully data-driven estimator attains the minimax rate in case **[o-o]** with $p > a$, **[o-s]** and **[s-o]** with $p \leqslant 1/2$, while in case **[o-o]** with $p \leqslant a$ and **[s-o]** with $p > 1/2$ the rate of the fully data-driven estimator $\widehat{\theta}^{(\eta)}$ features a deterioration by a logarithmic factor $(\log n_\lambda)^{p/a}$ and $(\log n)^{(2a+1)(1-1/(2p))}$, respectively, compared to the minimax rate.

3.5 Conclusion

We heave hence seen that the estimator we suggested in this chapter, which is motivated by the posterior mean of hierarchical Gaussian sieves, attains optimal oracle as well as minimax rates under mild assumptions. We have shown that the said assumptions can be verified for the inverse Gaussian sequence space model with known and unknown operator and in the circular density deconvolution model with known and unknown error density. We pointed out that in the case of absolutely regular processes, we need a strong hypothesis for the decay of the mixing coefficients which relaxation would be an interesting subject of study. It would also be interesting to study Gaussian processes defined on the entire real axis or deconvolution of densities of $\mathbb{R}$-valued random variables as it would require to investigate the case $\mathbb{F} = \mathbb{R}$.

Appendix A

Proof for SECTION 2.4.4

A.1 Proof of Theorem 2.4.2

A.1.1 Intermediate results

To prove this result, we will apply Theorem 2.3.3. Hence, we will verify Assumption 13. We will take the following expressions for the sequences G_n^+, and G_n^-.

DEFINITION 41 Define the following quantities :

$$\begin{aligned} G_n^- &:= \min\{m \in [\![1, m_n^\circ]\!] : \quad \mathfrak{b}_m^2(\theta^\circ) \leqslant 9L\Phi_n^\circ(\theta^\circ, \lambda)\}, \\ G_n^+ &:= \max\{m \in [\![m_n^\circ, G_n]\!] : n^{-1}(m - m_n^\circ) \leqslant 3\Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda)\}. \end{aligned}$$

□

With this choice, we have the following results, for which the proofs are given underneath.

PROPOSITION A.1.1.
Under ASSUMPTION 15*, we have, for all m in $[\![1, G_n]\!]$*

$$\begin{aligned} \mathbb{P}[\|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2 < [\mathfrak{b}_m^2(\theta^\circ) \vee n^{-1} m \Lambda_\circ(m)]/2] &\leqslant \exp[-m/(16L)], \\ \mathbb{P}[\|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2 > 4[\mathfrak{b}_m^2(\theta^\circ) \vee n^{-1} m \Lambda_\circ(m)]] &\leqslant \exp[-m/(9L)]. \end{aligned}$$

□

This first result implies the third condition of Assumption 13.

PROPOSITION A.1.2.
Under ASSUMPTION 15*, we have the following concentration inequalities for the threshold hyper parameter :*

$$\begin{aligned} \sum\nolimits_{m > G_n^+} \mathbb{P}(\Upsilon(m, \phi_n) - \Upsilon(m_n^\circ, \phi_n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)) &\leqslant \exp[-5m_n^\circ/(9L) + \log(G_n)]; \\ \sum\nolimits_{m < G_n^-} \mathbb{P}(\Upsilon(m, \phi_n) - \Upsilon(m_n^\circ, \phi_n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)) &\leqslant \exp[-4m_n^\circ/9 + \log(G_n)]. \end{aligned}$$

□

This second result implies the two remaining conditions of Assumption 13.
And we can hence directly apply Theorem 2.3.3 to obtain the considered theorem.

A.1.2 Detailed proofs

PROOF OF PROPOSITION A.1.1

Let be m in $[\![1, G_n]\!]$ and note that $\|\theta_{n,\overline{m}} - \theta^\circ\|^2 = \sum_{s=1}^{m} (\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s))^2 + \mathfrak{b}_m^2(\theta^\circ)$.

We will use LEMMA 2.4.1; therefor, define, for any s in $[\![1, m]\!]$

$$\begin{aligned}
S_m \quad &:= \sum\nolimits_{j=1}^{m} (\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s))^2; & \mu_m &:= \mathbb{E}[S_m] = n^{-1} m \Lambda_\circ(m);\\
\beta(s)^2 &:= \mathbb{V}[\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s)] = n^{-1}\Lambda(s); & v_m &:= \sum\nolimits_{j=1}^{m} \beta(s)^2 = n^{-1} m \Lambda_\circ(m);\\
\alpha(s)^2 &:= \mathbb{E}[\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s)] = 0; & t_m &:= \max_{j\in[\![1,m]\!]} \beta(s)^2 = n^{-1}\Lambda(m).
\end{aligned}$$

We then control the concentration of S_m, first from above, using that, for any a and b in $\mathbb{R}_+$, we have $a \vee b \leqslant a + b$; Lemma 2.4.1; and Assumption 15. We obtain

$$\begin{aligned}
&\mathbb{P}[S_m + \mathfrak{b}_m^2(\theta^\circ) \leqslant [n^{-1} m \Lambda_\circ(m) \vee \mathfrak{b}_m^2(\theta^\circ)]/2]\\
&\quad \leqslant \mathbb{P}[S_m + \mathfrak{b}_m^2(\theta^\circ) \leqslant (n^{-1} m \Lambda_\circ(m) + \mathfrak{b}_m^2(\theta^\circ))/2]\\
&\quad \leqslant \mathbb{P}[S_m + \mathfrak{b}_m^2(\theta^\circ) \leqslant (2n)^{-1} m \Lambda_\circ(m) + \mathfrak{b}_m^2(\theta^\circ)] \leqslant \mathbb{P}[S_m \leqslant (2n)^{-1} m \Lambda_\circ(m)]\\
&\quad \leqslant \mathbb{P}[S_m - \mu_m \leqslant -(2n)^{-1} m \Lambda_\circ(m)] \leqslant \exp[-(nm\Lambda_\circ(m))(16n\Lambda(m))^{-1}]
\end{aligned}$$
$\leqslant \exp[-m(16L)^{-1}]$.

Finally, we control the concentration of S_m from bellow using that, for any a and b in $\mathbb{R}_+$, we have $a \vee b \geqslant (a + b)/2$; Lemma 2.4.1; and Assumption 15; We obtain

$$\begin{aligned}
&\mathbb{P}[S_m + \mathfrak{b}_m^2(\theta^\circ) \geqslant 4[n^{-1} m \Lambda_\circ(m) \vee \mathfrak{b}_m^2(\theta^\circ)]]\\
&\quad \leqslant \mathbb{P}[S_m + \mathfrak{b}_m^2(\theta^\circ) \geqslant 2(n^{-1} m \Lambda_\circ(m) + \mathfrak{b}_m^2(\theta^\circ))]\\
&\quad \leqslant \mathbb{P}[S_m + \mathfrak{b}_m^2(\theta^\circ) \geqslant 2n^{-1} m \Lambda_\circ(m) + \mathfrak{b}_m^2(\theta^\circ)] \leqslant \mathbb{P}[S_m \geqslant 2n^{-1} m \Lambda_\circ(m)]\\
&\quad \leqslant \mathbb{P}[S_m - \mu_m \geqslant n^{-1} m \Lambda_\circ(m)] \leqslant \exp[-(nm\Lambda_\circ(m))(9n\Lambda(m))^{-1}] \leqslant \exp[-m(9L)^{-1}].
\end{aligned}$$

□

PROOF FOR PROPOSITION A.1.2

First, let's proof the first inequality. Use the fact that :

$$\begin{aligned}
&\sum\nolimits_{m > G_n^+} \mathbb{P}(\Upsilon(m, \phi_n) - \Upsilon(m_n^\circ, \phi_n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m))\\
&\quad = \sum\nolimits_{m=G_n^+ + 1}^{G_n} \mathbb{P}[0 < 3n^{-1}(m_n^\circ - m) + \sum\nolimits_{s=m_n^\circ+1}^{m} \phi_n(s)^2]
\end{aligned}$$

We will now use LEMMA 2.4.1. For this purpose, define then for all m in $[\![G_n^+ + 1, G_n]\!]$: $S_m := \sum_{j=m_n^\circ+1}^{m} \phi_n(s)^2$, we then have $\mu_m := \mathbb{E}[S_m] = n^{-1}(m - m_n^\circ) + \sum_{s=m_n^\circ+1}^{m} (\theta^\circ(s)\lambda(s))^2$, $\alpha(s)^2 := \mathbb{E}[\phi_n(s)]^2 = (\theta^\circ(s)\lambda(s))^2$ and $\beta(s)^2 := \mathbb{V}[\phi_n(s)] = n^{-1}$. Now, using that λ is monotonically decreasing and $\mathfrak{b}^2_{m_n^\circ}(\theta^\circ) \leqslant \Phi_n^\circ(\theta^\circ, \lambda)$, we note

$$\begin{aligned}\sum_{s=m_n^\circ+1}^{m} \alpha(s)^2 =& \sum_{s=m_n^\circ+1}^{m} (\theta^\circ(s)\lambda(s))^2 \leqslant \Lambda(m_n^\circ)^{-1} \sum_{j=m_n^\circ+1}^{m} (\theta^\circ(s))^2 \\ &\leqslant \Lambda(m_n^\circ)^{-1}\, \mathfrak{b}^2_{m_n^\circ}(\theta^\circ) \leqslant \Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda) =: r_m; \\ \sum_{s=m_n^\circ+1}^{m} \beta(s)^2 =& n^{-1}(m - m_n^\circ) =: v_m; \quad \max_{j\in[\![m_n^\circ+1, m]\!]} \beta_j = n^{-1} =: t_m\end{aligned}$$

Hence, we have, for all m in $[\![G_n^+, G_n]\!]$

$$\begin{aligned}&\mathbb{P}[\sum_{s=m_n^\circ+1}^{m} \phi_n(s)^2 - 3n^{-1}(m - m_n^\circ) > 0] \\ &= \mathbb{P}[S_m - n^{-1}(m - m_n^\circ) > 2n^{-1}(m - m_n^\circ)] \\ &\leqslant \mathbb{P}[S_m - \mu_m > 2n^{-1}(m - m_n^\circ) - \Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda)].\end{aligned}$$

Using the definition of G_n^+, we have $n^{-1}(m - m_n^\circ) > 3\Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda)$. Hence, we can write, using Assumption 15 and Lemma 2.4.1 with $c = 2/3$:

$$\begin{aligned}&\mathbb{P}[\sum_{s=m_n^\circ}^{m} \phi_n(s)^2 - 3n^{-1}(m - m_n^\circ) > 0] \\ &\leqslant \mathbb{P}[S_m - \mu_m > n^{-1}(m - m_n^\circ) + 2\Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda)] \\ &\leqslant \mathbb{P}[S_m - \mu_m > v_m + 2r_m] \leqslant \exp[-n(n^{-1}(m - m_n^\circ) + 2\Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda))/9] \\ &\leqslant \exp[-n(5\Lambda(m_n^\circ)^{-1}\Phi_n^\circ(\theta^\circ, \lambda))/9] \leqslant \exp[-5m_n^\circ/(9L)].\end{aligned}$$

Which gives

$$\sum_{m>G_n^+} \mathbb{P}(\Upsilon(m, \phi_n) - \Upsilon(m_n^\circ, \phi_n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)) \leqslant \exp[-5m_n^\circ/(9L) + \log(G_n)]$$

Hence, the hypothesis is verified.

We now prove the second inequality. We begin by noting that:

$$\begin{aligned}&\sum_{m<G_n^-} \mathbb{P}(\Upsilon(m, \phi_n) - \Upsilon(m_n^\circ, \phi_n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)) \\ &= \sum_{m=1}^{G_n^-} \mathbb{P}[\sum_{s=m+1}^{m_n^\circ} \phi_n(s)^2 < 3n^{-1}(m_n^\circ - m)].\end{aligned}$$

The Lemma 2.4.1 steps in again. Define $S_m := \sum_{s=m+1}^{m_n^\circ} \phi_n(s)^2$ and we want to control the concentration of this sum, hence we take the following notations :

$$\begin{aligned}\mu_m &:= \mathbb{E}[S_m] = n^{-1}(m_n^\circ - m) + \sum_{s=m+1}^{m_n^\circ} (\theta^\circ(s)\lambda(s))^2 \\ r_m &:= \sum_{j=m+1}^{m_n^\circ} (\theta^\circ(s)\lambda(s))^2; \quad v_m := n^{-1}(m_n^\circ - m); \quad t_m := n^{-1}.\end{aligned}$$

Hence, we have, using ASSUMPTION 15 and the definition of G_n^-

$$\begin{aligned}
&\mathbb{P}[S_m < 3n^{-1}(m_n^\circ - m)] \\
&\quad = \mathbb{P}[S_m - \mu_m < 3n^{-1}(m_n^\circ - m) - n^{-1}(m_n^\circ - m) - \sum\nolimits_{s=m+1}^{m_n^\circ}(\theta^\circ(s)\lambda(s))^2] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m < -[v_m + 2r_m]/3 + 7n^{-1}m_n^\circ/3 + \Lambda(m_n^\circ)^{-1}(\mathfrak{b}^2_{m_n^\circ}(\theta^\circ) - \mathfrak{b}^2_m(\theta^\circ))/3] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m < -[v_m + 2r_m]/3 + (1-2L)(\Phi_n^\circ(\theta^\circ,\lambda)\Lambda(m_n^\circ)^{-1})/3]
\end{aligned}$$

we now use LEMMA 2.4.1

$$\begin{aligned}
\mathbb{P}[S_m < 3n^{-1}(m_n^\circ - m)] &\leqslant \mathbb{P}[S_m - \mu_m < -[v_m + 2r_m]/3] \\
&\leqslant \exp[-n(n^{-1}(m_n^\circ - m) + 2\sum\nolimits_{s=m+1}^{m_n^\circ}(\theta^\circ(s)\lambda(s))^2)/36] \\
&\leqslant \exp[-n(16L\Phi_n^\circ(\theta^\circ,\lambda)\Lambda(m_n^\circ)^{-1})/36] \leqslant \exp[-4m_n^\circ/9].
\end{aligned}$$

And hence

$$\sum\nolimits_{m<G_n^-} \mathbb{P}(\Upsilon(m,\phi_n) - \Upsilon(m_n^\circ,\phi_n) < \mathrm{pen}(m_n^\circ) - \mathrm{pen}(m)) \leqslant \exp[-4m_n^\circ/9 + \log(G_n)].$$

□

A.2 Proof of Theorem 2.4.3

A.2.1 Intermediate results

Let us first consider the following intermediate results, for which the proofs are given later.

PROPOSITION A.2.1.
Under ASSUMPTION 15, we have, for all m in $[\![1, G_n]\!]$ and c greater than $3/2$,

$$\mathbb{P}[\|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2} > 4c[\mathfrak{b}^2_m(\theta^\circ) \vee n^{-1}(m\Lambda_\circ(m))]] \leqslant \exp[-cm(6L)^{-1}].$$

□

DEFINITION 42 Define the following quantities :

$$\begin{aligned}
G_n^{\star-} &:= \min\{m \in [\![1, m_n^\star]\!] : \quad \mathfrak{b}^2_m(\theta^\circ) \leqslant 9(1 \vee r)L\Phi_n^\star(\mathfrak{a},\lambda)\}, \\
G_n^{\star+} &:= \max\{m \in [\![m_n^\star, G_n]\!] : n^{-1}(m - m_n^\star) \leqslant 3\Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a},\lambda)\}.
\end{aligned}$$

□

PROPOSITION A.2.2.
Under ASSUMPTION 15, we have the following concentration inequalities for the threshold

hyper parameter :

$$\begin{aligned}\mathbb{P}[\widehat{m} > G_n^{\star+}] &\leqslant \exp[-(9L)^{-1}5(1 \vee r)m_n^\star + \log(G_n)],\\ \mathbb{P}[\widehat{m} < G_n^{\star-}] &\leqslant \exp[-7(1 \vee r)m_n^\star/9 + \log(G_n)].\end{aligned}$$

□

Then the proof of the theorem goes as follows.

Proof of Theorem 2.4.3

By the the total probability formula, we have :

$$\mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2 \leqslant K^\star\Phi_n^\star(\mathfrak{a},\lambda))] = 1 - \mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(K^\star\Phi_n^\star(\mathfrak{a},\lambda) < \|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2)].$$

Hence, we will control $\mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(K^\star\Phi_n^\star(\mathfrak{a},\lambda) < \|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2)]$.
We can write :

$$\begin{aligned}&\mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(K^\star\Phi_n^\star(\mathfrak{a},\lambda) < \|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2)]\\ &= \sum_{m=1}^{G_n} \mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(\{K^\star\Phi_n^\star(\mathfrak{a},\lambda) < \|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2\} \cap \{M = m\})]\\ &\leqslant \sum_{m=1}^{G_n^{\star-}-1} \mathbb{E}[\mathbb{P}^{(\infty)}_{M|\phi_n}(\{M = m\})] + \sum_{m=G_n^{\star+}+1}^{G_n} \mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n n}(\{M = m\})]\\ &+ \sum_{m=G_n^{\star-}}^{G_n^{\star+}} \mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n, M=m}(\{K^\star\Phi_n^\star(\mathfrak{a},\lambda) < \|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|_{l^2}^2\})]\\ &\leqslant \underbrace{\sum_{m=1}^{G_n^{\star-}-1} \mathbb{P}[\{\widehat{m} = m\}]}_{=:A} + \underbrace{\sum_{m=G_n^{\star+}+1}^{G_n} \mathbb{P}[\{\widehat{m} = m\}]}_{=:B}\\ &+ \sum_{m=G_n^{\star-}}^{G_n^{\star+}} \underbrace{\mathbb{P}[\{K^\star\Phi_n^\star(\mathfrak{a},\lambda) < \|\theta_{n,\overline{m}} - \theta^\circ\|_{l^2}^2\}]}_{=:C_m}.\end{aligned}$$

We control A and B using Proposition A.2.2. Hence, we now control $\sum_{m=G_n^{\star-}}^{G_n^{\star+}} C_m$. Using Assumption 17 we have

$$[a_{m_n^\star} \wedge n^{-1}m_n^\star\Lambda(m_n^\star)] \leqslant \Phi_n^\star(\mathfrak{a},\lambda) \leqslant (\kappa^\star)^{-1}[a_{m_n^\star} \wedge n^{-1}m_n^\star\Lambda(m_n^\star)].$$

Hence, for any m in $[\![m_n^\star, G_n^{\star+}]\!]$ we have, using the definition of $G_n^{\star+}$

$$\begin{aligned}m \leqslant& 3\Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a},\lambda)n + m_n^\star \leqslant 3(1 \vee r)n(\Lambda(m_n^\star)\kappa^\star)^{-1}[\mathfrak{a}_{m_n^\star} \wedge n^{-1}m_n^\star\Lambda_\circ(m_n^\star)] + m_n^\star\\ \leqslant& 3(1 \vee r)(\kappa^\star)^{-1}m_n^\star\Lambda_\circ(m_n^\star)\Lambda(m_n^\star)^{-1} + m_n^\star \leqslant (3(1 \vee r)(\kappa^\star L)^{-1} + 1)m_n^\star \leqslant D^\star m_n^\star;\end{aligned}$$

and $\Lambda_\circ(m) \leqslant \Lambda(m) \leqslant \Lambda(D^\star m_n^\star) \leqslant \Lambda(D^\star)\Lambda(m_n^\star) \leqslant \Lambda(D^\star)L\Lambda_\circ(m_n^\star)$; which give together

$$n^{-1}m\Lambda_\circ(m) \leqslant D^\star\Lambda(D^\star)Ln^{-1}m_n^\star\Lambda(m_n^\star) \leqslant D^\star\Lambda(D^\star)L\Phi_n^\star(\mathfrak{a},\lambda);$$

moreover, we have $\mathfrak{b}^2_m(\theta^\circ) \leqslant \mathfrak{b}^2_{m^\star_n}(\theta^\circ)^2(\theta^\circ) \leqslant \Phi^\star_n(\mathfrak{a}, \lambda)$, which leads to the conclusion

$$[\mathfrak{b}^2_m(\theta^\circ) \vee n^{-1} m \Lambda_\circ(m)] \leqslant D^\star \Lambda(D^\star)(1 \vee r)\Phi^\star_n(\mathfrak{a}, \lambda).$$

On the other hand, for and m in $[\![G^{\star-}_n, m^\star_n]\!]$, the definition of $G^{\star-}_n$ directly gives us

$$[\mathfrak{b}^2_m(\theta^\circ) \vee n^{-1} m \Lambda_\circ(m)] \leqslant 9L(1 \vee r)\Phi^\star_n(\mathfrak{a}, \lambda).$$

Using Proposition A.2.1, we have that, for all m in $[\![G^{\star-}_n, G^{\star+}_n]\!]$ and c greater than $3/2$:

$$\mathbb{P}[\{\|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2} > 4c(9L \vee D^\star\Lambda(D^\star))(1 \vee r)\Phi^\star_n(\mathfrak{a}, \lambda)\}] \leqslant \exp[-cm/(6L)].$$

Hence, we set $K^\star := 6(9L \vee D^\star \Lambda(D^\star))(1 \vee r)$, which leads us to the upper bound:

$$\sum\nolimits_{m=G^{\star-}_n}^{G^{\star+}_n} C_m \leqslant \sum\nolimits_{m=G^{\star-}_n}^{G^{\star+}_n} \exp[-3m/(8L)] \leqslant 4L \exp[-G^{\star-}_n(4L)].$$

Finally, we can conclude :

$$\begin{aligned}
&\mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|^2_{l^2} \leqslant K^\star \Phi^\star_n(\mathfrak{a}, \lambda))] \\
&\quad \geqslant 1 - \exp[-(5/(9L))(1 \vee r)m^\star_n + \log(G_n)] - \exp[-(7/9)(1 \vee r)m^\star_n + \log(G_n)] \\
&\quad - 4L \exp[-G^{\star-}_n/(4L)].
\end{aligned}$$

This proves the first part of the theorem for any θ° such that $G^{\star-}_n$ tends to infinity when n tends to ∞. In the opposite case, it means that there exist n° such that for all n larger than n°, $G^{\star-}_n = G^{\star-}_{n^\circ}$. This means that $n \mapsto \mathfrak{b}_{G^{\star-}_n}$ is constant function for n larger that n° but, by definition of $G^{\star-}_n$, we also have $\mathfrak{b}_{G^{\star-}_n} \leqslant 9(1 \vee r)L\Phi^\star_n(\mathfrak{a}, \lambda) \to 0$ which leads to the conclusion that for all m greater than $G^{\star-}_{n^\circ}$, $\mathfrak{b}^2_m(\theta^\circ) = 0$. Hence, for all m greater than $G^{\star-}_{n^\circ}$, we can write

$$\begin{aligned}
&K^\star \cdot \Phi^\star_n(\mathfrak{a}, \lambda)[\mathfrak{b}^2_m(\theta^\circ) \vee n^{-1} m \Lambda_\circ(m)]^{-1} = K^\star \Phi^\star_n(\mathfrak{a}, \lambda)[n^{-1} m \Lambda_\circ(m)]^{-1} \\
&\quad \geqslant K^\star n^{-1} m^\star_n \Lambda_\circ(m^\star_n)[n^{-1} m \Lambda_\circ(m)]^{-1} \geqslant K^\star m^\star_n (Lm)^{-1} \geqslant 9D^\star\Lambda(D^\star)(1 \vee r)m^{-1} m^\star_n \geqslant 1.
\end{aligned}$$

Hence, we set $c := \frac{9}{4} D^\star \Lambda(D^\star)(1 \vee r)\frac{m^\star_n}{m}$ and can write in this case for all n larger than n°:

$$\begin{aligned}
&\sum\nolimits_{m=G^{\star-}_n}^{G^{\star+}_n} \mathbb{P}[K^\star \Phi^\star_n(\mathfrak{a}, \lambda) < \|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2}] \\
&\quad \leqslant \sum\nolimits_{m=G^{\star-}_n}^{G^{\star+}_n} \mathbb{P}[4c[\mathfrak{b}^2_m(\theta^\circ) \vee n^{-1} m \Lambda_\circ(m)] < \|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2}] \\
&\quad \leqslant \sum\nolimits_{m=G^{\star-}_n}^{G^{\star+}_n} \exp[-cm/(6L)] \leqslant \sum\nolimits_{m=G^{\star-}_n}^{G^{\star+}_n} \exp[-3/(8L)D^\star\Lambda(D^\star)(1 \vee r)m^\star_n] \\
&\quad \leqslant \exp[-3/(8L)D^\star\Lambda(D^\star)(1 \vee r)m^\star_n + \log(G_n)].
\end{aligned}$$

We can hence conclude that

$$\begin{aligned}&\mathbb{E}[\mathbb{P}^{(\infty)}_{\boldsymbol{\theta}_{\overline{M}}|\phi_n}(\|\boldsymbol{\theta}_{\overline{M}} - \theta^\circ\|^2_{l^2} \leqslant K^\star\Phi^\star_n(\mathfrak{a},\lambda))]\\ &\quad > 1 - \exp[-3/(8L)D^\star\Lambda(D^\star)(1\vee r)m^\star_n + \log(G_n)]\\ &\quad - \exp[-5m^\star_n/(9L) + \log(G_n)] - \exp[-(7/9)m^\star_n + \log(G_n)].\end{aligned}$$

Hence, we have shown here that $\Phi^\star_n(\mathfrak{a},\lambda)$ is an upper bound for the contraction rate under the quadratic risk. We will now use this to prove that it is also for the maximal risk. Note that $K^\star\,\Phi^\star_n(\mathfrak{a},\lambda) \geqslant 4[\mathfrak{b}^2_m(\theta^\circ)\vee n^{-1}m\Lambda_\circ(m)]$ for all m in $[\![G^{\star-}_n, G^{\star+}_n]\!]$. Hence, for any increasing function K_n such that $\lim_{n\to\infty} K_n = \infty$, we have

$$K_n\Phi^\star_n(\mathfrak{a},\lambda) \geqslant 4K_n(K^\star)^{-1}[\mathfrak{b}^2_m(\theta^\circ)\vee n^{-1}m\Lambda_\circ(m)].$$

So, if we define $\tilde{n}^\circ$, the smallest integer such that $K_n(K^\star)^{-1}\geqslant 1$, we can apply Proposition A.2.1 and we have:

$$\begin{aligned}&\sum_{m=G^{\star-}_n}^{G^{\star+}_n} \mathbb{P}[K_n\Phi^\star_n(\mathfrak{a},\lambda) < \|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2}]\\ &\quad\leqslant \sum_{m=G^{\star-}_n}^{G^{\star+}_n} \mathbb{P}[4K_n(K^\star)^{-1}[\mathfrak{b}^2_m(\theta^\circ)\vee n^{-1}m\Lambda_\circ(m)] < \|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2}]\\ &\quad\leqslant \sum_{m=G^{\star-}_n}^{G^{\star+}_n} \exp[-4K_n m(9K^\star L)^{-1}] \leqslant \exp[-4K_n(9K^\star L)^{-1}].\end{aligned}$$

We hence here have a uniform upper bound for the maximal risk which concludes the proof.
□

A.2.2 Detailed proofs

Proof of Proposition A.2.1

Let be m in $[\![1, G_n]\!]$ and note that $\|\theta_{n,\overline{m}} - \theta^\circ\|^2_{l^2} = \sum_{s=1}^m(\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s))^2 + \mathfrak{b}^2_m(\theta^\circ)$, hence, we will use Lemma 2.4.1. We then define for any s in $[\![1, m]\!]$

$$\begin{aligned}&S_m := \sum_{s=1}^m(\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s))^2; && \mu_m := \mathbb{E}[S_m] = n^{-1}m\Lambda_\circ(m),\\ &\beta(s)^2 := \mathbb{V}[\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s)] = n^{-1}\Lambda(s); && v_m := \sum_{s=1}^m\beta(s)^2 = n^{-1}m\Lambda_\circ(m);\\ &\alpha(s)^2 := \mathbb{E}[\phi_n(s)\lambda(s)^{-1} - \theta^\circ(s)] = 0; && t_m := \max_{s\in[\![1,m]\!]}\beta(s)^2 = n^{-1}\Lambda(m).\end{aligned}$$

We then control the concentration of S_m, define c a constant greater than $3/2$. Using that, for any a and b in $\mathbb{R}_+$ we have $a\vee b \geqslant (a+b)/2$; $c > 1$; and Lemma 2.4.1, we obtain

$$\begin{aligned}&\mathbb{P}[S_m + \mathfrak{b}^2_m(\theta^\circ) \geqslant 4c(n^{-1}m\Lambda_\circ(m)\vee\mathfrak{b}^2_m(\theta^\circ))]\\ &\quad\leqslant \mathbb{P}[S_m + \mathfrak{b}^2_m(\theta^\circ) \geqslant 2cn^{-1}m\Lambda_\circ(m) + \mathfrak{b}^2_m(\theta^\circ)] \leqslant \mathbb{P}[S_m \geqslant 2cn^{-1}m\Lambda_\circ(m)]\\ &\quad\leqslant \mathbb{P}[S_m - \mu_m \geqslant cn^{-1}m\Lambda_\circ(m)] \leqslant \exp[-cnm\Lambda_\circ(m)(6n\Lambda(m))^{-1}] \leqslant \exp[-cm/(6L)];\end{aligned}$$

which proves the claim □

Proof of Proposition A.2.2
First, let's proof the first inequality. Use the fact that :

$$\begin{aligned}
&\mathbb{P}[G_n^{\star+} < \widehat{m} \leqslant G_n] \\
&\quad = \mathbb{P}[\forall l \in [\![1, G_n^{\star+}]\!], \quad n^{-1}3\widehat{m} - \sum\nolimits_{s=1}^{\widehat{m}} \phi_n(s)^2 < n^{-1}3l - \sum\nolimits_{s=1}^{l} \phi_n(s)^2] \\
&\quad \leqslant \mathbb{P}[\exists m \in [\![G_n^{\star+}+1, G_n]\!] : \quad n^{-1}3m - \sum\nolimits_{s=1}^{m} \phi_n(s)^2 < n^{-1}3m_n^\star - \sum\nolimits_{s=1}^{m_n^\star} \phi_n(s)^2] \\
&\quad \leqslant \sum\nolimits_{m=G_n^{\star+}+1}^{G_n} \mathbb{P}[n^{-1}3m - \sum\nolimits_{s=1}^{m} \phi_n(s)^2 < n^{-1}3m_n^\star - \sum\nolimits_{s=1}^{m_n^\star} \phi_n(s)^2] \\
&\quad \leqslant \sum\nolimits_{m=G_n^{\star+}+1}^{G_n} \mathbb{P}[0 < 3n^{-1}m_n^\star - m + \sum\nolimits_{s=m_n^\star+1}^{m} \phi_n(s)^2]
\end{aligned}$$

We will now use Lemma 2.4.1. For this purpose, define then for all m in $[\![G_n^{\star+}+1, G_n]\!]$: $S_m := \sum_{s=m_n^\star+1}^{m} \phi_n(s)^2$, we then have $\mu_m := \mathbb{E}[S_m] = n^{-1}m - m_n^\star + \sum_{s=m_n^\star+1}^{m} \phi(s)^2$, $\alpha(s)^2 := \mathbb{E}[\phi_n(s)]^2 = \phi(s)^2$ and $\beta(s)^2 := \mathbb{V}[\phi_n(s)] = n^{-1}$. Now we note, using the definition of $\Theta(\mathfrak{a}, r)$

$$\begin{aligned}
\sum\nolimits_{s=m_n^\star+1}^{m} \alpha(s)^2 &= \sum\nolimits_{s=m_n^\star+1}^{m} \phi(s)^2 \leqslant \Lambda(m_n^\star)^{-1} \sum\nolimits_{s=m_n^\star+1}^{m} (\theta^\circ(s))^2 \\
&\leqslant \Lambda(m_n^\star)^{-1}\, \mathfrak{b}^2_{m_n^\star}(\theta^\circ) \leqslant \Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda) \qquad =: r_m; \\
\sum\nolimits_{s=m_n^\star+1}^{m} \beta(s)^2 &= n^{-1}(m - m_n^\star) =: v_m; \quad \max_{j \in [\![m_n^\star, m]\!]} \beta(s) = n^{-1} =: t_m.
\end{aligned}$$

Hence, we have, for all m in $[\![G_n^{\star+}+1, G_n]\!]$

$$\begin{aligned}
&\mathbb{P}[\sum\nolimits_{s=m_n^\star+1}^{m} \phi_n(s)^2 - 3n^{-1}(m - m_n^\star) > 0] = \mathbb{P}[S_m - n^{-1}(m - m_n^\star) > 2n^{-1}(m - m_n^\star)] \\
&\quad = \mathbb{P}[S_m - n^{-1}(m - m_n^\star) - \sum\nolimits_{s=m_n^\star+1}^{m} \phi(s)^2 > 2n^{-1}(m - m_n^\star) - \sum\nolimits_{s=m_n^\star+1}^{m} \phi(s)^2] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m > 2n^{-1}(m - m_n^\star) - \Lambda(m_n^\star)^{-1}\, \mathfrak{b}^2_{m_n^\star}(\theta^\circ)] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m > 2n^{-1}(m - m_n^\star) - \Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda)].
\end{aligned}$$

Using the definition of $G_n^{\star+}$, we have $n^{-1}(m - m_n^\star) > 3\Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda)$. Hence, we can write :

$$\begin{aligned}
&\mathbb{P}[\sum\nolimits_{s=m_n^\star}^{m} \phi_n(s)^2 - 3n^{-1}(m - m_n^\star) > 0] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m > n^{-1}(m - m_n^\star) + 2\Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda)] \leqslant \mathbb{P}[S_m - \mu_m > v_m + 2r_m] \\
&\quad \leqslant \exp[-n(n^{-1}(m - m_n^\star) + 2\Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda))/9] \\
&\quad \leqslant \exp[-n(5\Lambda(m_n^\star)^{-1}(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda))/9] \leqslant \exp[-5(1 \vee r)m_n^\star/(9L)].
\end{aligned}$$

Finally we can conclude that

$$\mathbb{P}[G_n^{\star+} < \widehat{m} \leqslant G_n] \leqslant \exp[-5(1 \vee r)m_n^\star/(9L) + \log(G_n)].$$

We now prove the second inequality. We begin by writing the same kind of inclusion of events as for the first inequality :

$$\begin{aligned}
&\mathbb{P}[1 \leqslant \widehat{m} < G_n^{\star -}] \\
&\quad = \mathbb{P}[\forall m \in [\![G_n^-, G_n]\!], \quad 3n^{-1}\widehat{m} - \sum_{s=1}^{\widehat{m}} \phi_n(s)^2 < 3n^{-1}m - \sum_{s=1}^{m} \phi_n(s)^2] \\
&\quad \leqslant \mathbb{P}[\exists m \in [\![1, G_n^{\star -} - 1]\!], \quad 3n^{-1}m - \sum_{s=1}^{m} \phi_n(s)^2 < 3n^{-1}m_n^\star - \sum_{s=1}^{m_n^\star} \phi_n(s)^2] \\
&\quad \leqslant \sum_{m=1}^{G_n^{\star -}} \mathbb{P}[3n^{-1}m - \sum_{s=1}^{m} \phi_n(s)^2 < 3n^{-1}m_n^\star - \sum_{s=1}^{m_n^\star} \phi_n(s)^2] \\
&\quad \leqslant \sum_{m=1}^{G_n^{\star -}} \mathbb{P}[\sum_{s=m+1}^{m_n^\star} \phi_n(s)^2 < 3n^{-1}(m_n^\star - m)].
\end{aligned}$$

The Lemma 2.4.1 steps in again, define $S_m := \sum_{s=m+1}^{m_n^\star} \phi_n(s)^2$ and we want to control the concentration of this sum, hence we take the following notations :

$$\begin{aligned}
\mu_m &:= \mathbb{E}[S_m] = n^{-1}(m_n^\star - m) + \sum_{s=m+1}^{m_n^\star} \phi(s)^2; \\
r_m &:= \sum_{s=m+1}^{m_n^\star} \phi(s)^2; \quad v_m := n^{-1}(m_n^\star - m); \quad t_m := n^{-1}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\mathbb{P}[S_m < 3n^{-1}(m_n^\star - m)] \\
&\quad = \mathbb{P}[S_m - \mu_m < 3n^{-1}(m_n^\star - m) - n^{-1}(m_n^\star - m) - \sum_{s=m+1}^{m_n^\star} \phi(s)^2] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m < (7/3)n^{-1}(m_n^\star - m) - (1/3)\Lambda(m_n^\star)^{-1} \sum_{s=m+1}^{m_n^\star} (\theta^\circ(s))^2 - [v_m + 2r_m]/3] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m < -[v_m + 2r_m]/3 + 3(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda)\Lambda_\circ(m_n^\star)^{-1} - (1/3)\Lambda(m_n^\star)^{-1}\, \mathfrak{b}_m^2(\theta^\circ)]
\end{aligned}$$

now, using the definition of G_n^-, we have $\mathfrak{b}_m^2(\theta^\circ) > 9L(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda)$ so

$$\begin{aligned}
&\mathbb{P}[S_m < 3n^{-1}(m_n^\star - m)] \\
&\quad \leqslant \mathbb{P}[S_m - \mu_m < -(1/3)[v_m + 2r_m]] \\
&\quad \leqslant \exp[-(n/36)(n^{-1}(m_n^\star - m) + 2\sum_{s=m+1}^{m_n^\star} \phi(s)^2)] \\
&\quad \leqslant \exp[-(n/36)(n^{-1}(m_n^\star - m) + 2\Lambda(m_n^\star)^{-1}\, \mathfrak{b}_m^2(\theta^\circ) - 2\Lambda(m_n^\star)^{-1}\, \mathfrak{b}_{m_n^\star}^2(\theta^\circ))] \\
&\quad \leqslant \exp[-(n/36)(16L(1 \vee r)\Phi_n^\star(\mathfrak{a}, \lambda)\Lambda(m_n^\star)^{-1})] \\
&\quad \leqslant \exp[-(4/9)(1 \vee r)m_n^\star]
\end{aligned}$$

Which in turn implies

$$\mathbb{P}[1 \leqslant \widehat{m} < G_n^{\star -}] \leqslant \exp[-(4/9)(1 \vee r)m_n^\star + \log(G_n)].$$

□

Appendix B

Proof for section 3.2

Proof of Lemma 3.2.1.
Consider (i). For $l > k$ remind that we have $\Pi_{\underline{l}\overline{k}} = \Pi_{\overline{l}} - \Pi_{\overline{k}} = \Pi_{\overline{k}}^{\perp} - \Pi_{\overline{l}}^{\perp}$ the orthogonal projection onto $\Theta_{\underline{k},\overline{l}} := \overline{lin}\{(\mathbb{1}_{s'=s})_{s'\in\mathbb{Z}}, |s| \in \rrbracket k, l\rrbracket\}$ where $\|\Pi_{\underline{l}\overline{k}}[x]\|_{l^2}^2 = \|\Pi_{\overline{l}}[x]\|_{l^2}^2 - \|\Pi_{\overline{k}}[x]\|_{l^2}^2 = \|\Pi_{\overline{k}}^{\perp}[x]\|_{l^2}^2 - \|\Pi_{\overline{l}}^{\perp}[x]\|_{l^2}^2$ for all $[x] \in \mathcal{L}^2$. Let us define $\gamma([x]) := \|[x]\|_{l^2}^2 - 2\langle [x]|\breve{\theta}_{\cdot}\rangle_{l^2}$. For each $l \in \llbracket 1, n\rrbracket$ and for any $[x] \in \Theta_{\overline{l}}$ we have $\langle [x]|\breve{\theta}_{\cdot}\rangle_{l^2} = \langle [x]|\breve{\theta}_{\overline{l}}\rangle_{l^2}$, which in turn implies $\gamma([x]) = \|[x]\|_{l^2}^2 - 2\langle [x]|\breve{\theta}_{\overline{l}}\rangle_{l^2} + \|\breve{\theta}_{\overline{l}}\|_{l^2}^2 - \|\breve{\theta}_{\overline{l}}\|_{l^2}^2 = \|[x] - \breve{\theta}_{\overline{l}}\|_{l^2}^2 - \|\breve{\theta}_{\overline{l}}\|_{l^2}^2$ and consequently, $\gamma(\breve{\theta}_{\overline{l}}) = -\|\breve{\theta}_{\overline{l}}\|_{l^2}^2$ for all $l \in \mathbb{N}$. Obviously, $\|\breve{\theta}_{\overline{k}}\|_{l^2}^2 - \|\breve{\theta}_{\overline{l}}\|_{l^2}^2 = \gamma(\breve{\theta}_{\overline{l}}) - \gamma(\breve{\theta}_{\overline{k}})$, while $\|\Pi_{\underline{l}\overline{k}}\theta^{\circ}\|_{l^2}^2 = \|\theta^{\circ}_{\underline{0}}\|_{l^2}^2\{\mathfrak{b}_k^2(\theta^{\circ}) - \mathfrak{b}_l^2(\theta^{\circ})\}$. Consequently, the claim (i) can equivalently be rewritten as

$$\gamma(\breve{\theta}_{\overline{l}}) - \gamma(\breve{\theta}_{\overline{k}}) \leqslant \tfrac{11}{2}\|\breve{\theta}_{\overline{l}} - \theta^{\circ}_{\overline{l}}\|_{l^2}^2 - \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}}\theta^{\circ}\|_{l^2}^2. \tag{B.1}$$

Analogously, if $k > l$ then the claim (ii) can equivalently be rewritten as

$$\gamma(\breve{\theta}_{\overline{l}}) - \gamma(\breve{\theta}_{\overline{k}}) \leqslant \tfrac{7}{2}\|\breve{\theta}_{\overline{k}} - \theta^{\circ}_{\overline{k}}\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{k}\overline{l}}\theta^{\circ}\|_{l^2}^2. \tag{B.2}$$

Proof of (B.1). For $x, y, z \in l_2$ we observe that

$$\begin{aligned}\gamma(x) - \gamma(y) &= \|x\|_{l^2}^2 - \|y\|_{l^2}^2 - 2\langle x - y|\breve{\theta}_{\cdot}\rangle_{l^2} \\ &= \|x\|_{l^2}^2 - 2\langle x|z\rangle_{l^2} + \|z\|_{l^2}^2 - \|y\|_{l^2}^2 + 2\langle y|z\rangle_{l^2} - \|z\|_{l^2}^2 - 2\langle x - y|\breve{\theta}_{\cdot} - z\rangle_{l^2} \\ &= \|x - z\|_{l^2}^2 - \|y - z\|_{l^2}^2 - 2\langle x - y|\breve{\theta}_{\cdot} - z\rangle_{l^2}\end{aligned}$$

and in particular for $x_{\overline{l}} = \Pi_{\overline{l}}x$ and $x_{\overline{k}} = \Pi_{\overline{k}}x$ with $k < l$ where $x_{\overline{l}} - x_{\overline{k}} = \Pi_{\underline{l}\overline{k}}x$ we have

$$\begin{aligned}\gamma(x_{\overline{l}}) - \gamma(x_{\overline{k}}) &= \|x_{\overline{l}} - z\|_{l^2}^2 - \|x_{\overline{k}} - z\|_{l^2}^2 - 2\langle x_{\overline{l}} - x_{\overline{k}}|\breve{\theta}_{\cdot} - z\rangle_{l^2} \\ &= \|\Pi_{\overline{l}}(x - z)\|_{l^2}^2 - \|\Pi_{\overline{k}}(x - z)\|_{l^2}^2 + \|\Pi_{\overline{l}}^{\perp}z\|_{l^2}^2 - \|\Pi_{\overline{k}}^{\perp}z\|_{l^2}^2 - 2\langle \Pi_{\underline{l}\overline{k}}x|\breve{\theta}_{\cdot} - z\rangle_{l^2} \\ &= \|\Pi_{\underline{l}\overline{k}}(x - z)\|_{l^2}^2 - \|\Pi_{\underline{l}\overline{k}}z\|_{l^2}^2 - 2\langle \Pi_{\underline{l}\overline{k}}x|\Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - z)\rangle_{l^2} \\ &= \|\Pi_{\underline{l}\overline{k}}(x - z)\|_{l^2}^2 - \|\Pi_{\underline{l}\overline{k}}z\|_{l^2}^2 - 2\|\Pi_{\underline{l}\overline{k}}x\|_{l^2}\langle \tfrac{\Pi_{\underline{l}\overline{k}}x}{\|\Pi_{\underline{l}\overline{k}}x\|_{l^2}}|\Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - z)\rangle_{l^2}.\end{aligned} \tag{B.3}$$

Exploiting the elementary inequality $-2ab \leqslant \frac{1}{4}a^2 + 4b^2$ and setting $\mathbb{B}_{kl} := \{x \in \Theta_{\underline{k},\overline{l}} :$

$\|x\|_{l^2} = 1\}$ it follows

$$\begin{aligned}
\gamma(x_l) - \gamma(x_k) &\leqslant \|\Pi_{\underline{l}\overline{k}}(x-z)\|_{l^2}^2 - \|\Pi_{\underline{l}\overline{k}} z\|_{l^2}^2 + \tfrac{1}{4}\|\Pi_{\underline{l}\overline{k}} x\|_{l^2}^2 + 4|\langle \tfrac{\Pi_{\underline{l}\overline{k}} x}{\|\Pi_{\underline{l}\overline{k}} x\|_{l^2}} | \Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - z)\rangle_{l^2}|^2 \\
&\leqslant \|\Pi_{\underline{l}\overline{k}}(x-z)\|_{l^2}^2 - \|\Pi_{\underline{l}\overline{k}} z\|_{l^2}^2 + \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}}(x-z)\|_{l^2}^2 + \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}} z\|_{l^2}^2 + 4 \sup_{y\in\mathbb{B}_{kl}} |\langle y | \Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - z)\rangle_{l^2}|^2 \\
&= \tfrac{3}{2}\|\Pi_{\underline{l}\overline{k}}(x-z)\|_{l^2}^2 - \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}} z\|_{l^2}^2 + 4\|\Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - z)\|_{l^2}^2.
\end{aligned}$$

Replacing x by $\breve{\theta}_{\cdot}$ and z by θ° the last estimate implies (B.1), that is,

$$\begin{aligned}
\gamma(\breve{\theta}_{\overline{l}}) - \gamma(\breve{\theta}_{\overline{k}}) &\leqslant \tfrac{3}{2}\|\Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - \theta^\circ)\|_{l^2}^2 - \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}}\theta^\circ\|_{l^2}^2 + 4\|\Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - \theta^\circ)\|_{l^2}^2 \\
&= \tfrac{11}{2}\|\Pi_{\underline{l}\overline{k}}(\breve{\theta}_{\cdot} - \theta^\circ)\|_{l^2}^2 - \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}}\theta^\circ\|_{l^2}^2 \leqslant \tfrac{11}{2}\|\breve{\theta}_{\overline{l}} - \theta^\circ_{\overline{l}}\|_{l^2}^2 - \tfrac{1}{2}\|\Pi_{\underline{l}\overline{k}}\theta^\circ\|_{l^2}^2
\end{aligned}$$

Proof of (B.2). In case of $k > l$ from (B.3) follows

$$\begin{aligned}
\gamma(x_{\overline{l}}) - \gamma(x_{\overline{k}}) &= -(\gamma(x_{\overline{k}}) - \gamma(x_{\overline{l}})) \\
&= -\|\Pi_{\underline{k}\overline{l}}(x-z)\|_{l^2}^2 + \|\Pi_{\underline{k}\overline{l}} z\|_{l^2}^2 + 2\|\Pi_{\underline{k}\overline{l}} x\|_{l^2}\langle \tfrac{\Pi_{\underline{k}\overline{l}} x}{\|\Pi_{\underline{k}\overline{l}} x\|_{l^2}} | \Pi_{\underline{k}\overline{l}}(\breve{\theta}_{\cdot} - z)\rangle_{l^2}
\end{aligned}$$

Exploiting again the elementary inequality $-2ab \leqslant \frac{1}{4}a^2 + 4b^2$ and keeping in mind that $\mathbb{B}_{lk} := \{x \in \Theta_{\underline{l},\overline{k}} : \|x\|_{l^2} = 1\}$ it follows

$$\begin{aligned}
\gamma(x_{\overline{l}}) - \gamma(x_{\overline{k}}) &\leqslant -\|\Pi_{\underline{k}\overline{l}}(x-z)\|_{l^2}^2 + \|\Pi_{\underline{k}\overline{l}} z\|_{l^2}^2 + \tfrac{1}{2}\|\Pi_{\underline{k}\overline{l}}(x-z)\|_{l^2}^2 \\
&\qquad + \tfrac{1}{2}\|\Pi_{\underline{k}\overline{l}} z\|_{l^2}^2 + 4 \sup_{y\in\mathbb{B}_{lk}} |\langle y | \Pi_{\underline{k}\overline{l}}(\breve{\theta}_{\cdot} - z)\rangle_{l^2}|^2 \\
&= -\tfrac{1}{2}\|\Pi_{\underline{k}\overline{l}}(x-z)\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{k}\overline{l}} z\|_{l^2}^2 + 4\|\Pi_{\underline{k}\overline{l}}(\breve{\theta}_{\cdot} - z)\|_{l^2}^2.
\end{aligned}$$

Replacing x by $\breve{\theta}_{\cdot}$ and z by θ° the last estimate implies (B.1), that is

$$\begin{aligned}
\gamma(\breve{\theta}_{\overline{l}}) - \gamma(\breve{\theta}_{\overline{k}}) &\leqslant -\tfrac{1}{2}\|\Pi_{\underline{k}\overline{l}}(\breve{\theta}_{\cdot} - \theta^\circ)\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{k}\overline{l}}\theta^\circ\|_{l^2}^2 + 4\|\Pi_{\underline{k}\overline{l}}(\breve{\theta}_{\cdot} - \theta^\circ)\|_{l^2}^2 \\
&= \tfrac{7}{2}\|\Pi_{\underline{k}\overline{l}}(\breve{\theta}_{\cdot} - \theta^\circ)\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{k}\overline{l}}\theta^\circ\|_{l^2}^2 \leqslant \tfrac{7}{2}\|\breve{\theta}_{\overline{k}} - \theta^\circ_{\overline{k}}\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{k}\overline{l}}\theta^\circ\|_{l^2}^2,
\end{aligned}$$

which completes the proof. □

B.1 Proofs for section 3.2.1

We present here the results in specific case of $\mathbb{F} = \mathbb{N}$ or $\mathbb{Z}$ and $\mathbb{M} = \mathbb{N}$, note, though, that they could be easily generalised to any of nested sieve.

PROOF OF LEMMA 3.2.2.

We start the proof with the observation that $\theta_n(s) - \theta^\circ(s) = \lambda^{-1}(s)(\phi_n(s) - \phi(s))\,\mathbb{P}_M^{(\eta)}([\![|s|, n]\!]) - \theta^\circ(s)\,\mathbb{P}_M^{(\eta)}([\![1, |s|[\![)$ for all s with $|s|$ in $[\![0, n]\!]$ and $\theta_n(s) - \theta^\circ(s) = -\theta^\circ(s)$ for all s with $|s| > n$. Consequently, (keep in mind that

$|\lambda^{-1}(s)|^2 = \Lambda(s)$) we have

$$\begin{aligned}
&\|\theta_n - \theta^\circ\|_{l^2}^2 = \\
&\sum_{|s|\in[\![0,n]\!]} |\lambda^{-1}(s)(\phi_n(s) - \phi(s))\,\mathbb{P}_M^{(\eta)}([\![|s|,n]\!]) - \theta^\circ(s)\,\mathbb{P}_M^{(\eta)}([\![1,|s|[\![)|^2 + \sum_{|s|>n} |\theta^\circ(s)|^2 \\
&\qquad\leqslant \sum_{|s|\in[\![1,n]\!]} 2\{\Lambda(s)|\phi_n(s) - \phi(s)|^2\,\mathbb{P}_M^{(\eta)}([\![|s|,n]\!])\} \\
&\qquad\quad + \sum_{|s|\in[\![1,n]\!]} 2|\theta^\circ(s)|^2\,\mathbb{P}_M^{(\eta)}([\![1,|s|[\![) + 2\sum_{|s|>n} |\theta^\circ(s)|^2, \qquad \text{(B.4)}
\end{aligned}$$

where we consider the first r.h.s and the two other r.h.s. terms separately. Consider the first r.h.s. term in (B.4). We split the sum into two parts which we bound separately. Precisely, for any m_+ in $[\![1,n]\!]$, we have

$$\begin{aligned}
&2\sum_{|s|\in[\![0,n]\!]} \Lambda(s)|\phi_n(s) - \phi(s)|^2\,\mathbb{P}_M^{(\eta)}([\![|s|,n]\!]) \\
&\qquad\leqslant \|\theta_{n,\overline{m_+}} - \theta^\circ_{\overline{m_+}}\|_{l^2}^2 + \sum_{|l|\in]\!]m_+,n]\!]} \mathbb{P}_M^{(\eta)}(l)\|\theta_{n,\bar{l}} - \theta^\circ_{\bar{l}}\|_{l^2}^2 \\
&\qquad\leqslant \|\theta_{n,\overline{m_+}} - \theta^\circ_{\overline{m_+}}\|_{l^2}^2 + \sum_{|l|\in]\!]m_+,n]\!]} \mathbb{P}_M^{(\eta)}(l)\big(\|\theta_{n,\bar{l}} - \theta^\circ_{\bar{l}}\|_{l^2}^2 - \mathrm{pen}(l)/7\big)_+ \\
&+\tfrac{1}{7}\sum_{|l|\in]\!]m_+,n]\!]} \mathbb{P}_M^{(\eta)}(l)\,\mathrm{pen}(l)\mathbb{1}_{\{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|_{l^2}^2 \geqslant \mathrm{pen}(l)/7\}} + \tfrac{1}{7}\sum_{l\in]\!]m_+,n]\!]} \mathrm{pen}(l)\,\mathbb{P}_M^{(\eta)}(l)\mathbb{1}_{\{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|_{l^2}^2 < \mathrm{pen}(l)/7\}} \\
&\qquad\leqslant \tfrac{1}{7}\,\mathrm{pen}(m_+) + \sum_{|l|\in[\![m_+,n]\!]} \big(\|\theta_{n,\bar{l}} - \theta^\circ_{\bar{l}}\|_{l^2}^2 - \mathrm{pen}(l)/7\big)_+ \\
&\qquad\quad + \tfrac{1}{7}\sum_{|l|\in]\!]m_+,n]\!]} \mathbb{P}_M^{(\eta)}(l)\,\mathrm{pen}(l)\mathbb{1}_{\{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|_{l^2}^2 \geqslant \mathrm{pen}(l)/7\}} \\
&\qquad\quad + \tfrac{1}{7}\sum_{|l|\in]\!]m_+,n]\!]} \mathrm{pen}(l)\,\mathbb{P}_M^{(\eta)}(l)\mathbb{1}_{\{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|_{l^2}^2 < \mathrm{pen}(l)/7\}} \qquad \text{(B.5)}
\end{aligned}$$

Consider the second and third r.h.s. term in (B.4). Splitting the first sum into two parts we obtain, for any m_+ in $[\![1,n]\!]$,

$$\begin{aligned}
&2\sum_{|s|\in[\![0,n]\!]} |\theta^\circ(s)|^2\,\mathbb{P}_M^{(\eta)}([\![0,|s|[\![) + 2\sum_{|s|>n} |\theta^\circ(s)|^2 \\
&\qquad\leqslant 2\sum_{|s|\in[\![0,m_-]\!]} |\theta^\circ(s)|^2\,\mathbb{P}_M^{(\eta)}([\![0,s[\![) + 2\sum_{|s|\in]\!]m_-,n]\!]} |\theta^\circ(s)|^2 + 2\sum_{|s|>n} |\theta^\circ(s)|^2 \\
&\qquad\qquad\leqslant \|\theta^\circ_{\underline{0}}\|_{l^2}^2\{\mathbb{P}_M^{(\eta)}([\![0,m_-[\![) + \mathfrak{b}_{m_-}^2(\theta^\circ)\} \qquad \text{(B.6)}
\end{aligned}$$

Combining (B.4) and the upper bounds (B.5) and (B.6) we obtain the assertion, which completes the proof. □

B.1.1 Proofs for section 3.2.1.1

Bellow we state the proof of Lemma 3.2.3 based on Lemma B.1.1, stated first.

Lemma B.1.1.
Consider the data-driven aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3). Using the weights as specified in eq. (3.4), for any $l \in [\![1,n]\!]$ with $\mathfrak{R}_n^l := \mathfrak{R}_n^l(\theta^\circ,\Lambda)$ holds

(i) *for all $k \in [\![1,l[\![$ we have*
$$\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|^2_{l^2}<\kappa\mathfrak{R}_n^l/7\}} \leqslant \exp\big(\eta n\{-\tfrac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}}{2}\,\mathfrak{b}_m^2(\theta^\circ)+[\tfrac{25\kappa}{14}+\tfrac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}}{2}]\mathfrak{R}_n^l-\mathrm{pen}^\Lambda(m)\}\big)$$

(ii) *for all $m \in]\!]l,n]\!]$ we have*
$$\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2}<\mathrm{pen}^\Lambda(m)/7\}} \leqslant \exp\big(\eta n\{-\tfrac{1}{2}\,\mathrm{pen}^\Lambda(m)+[\tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}+\kappa]\mathfrak{R}_n^l\}\big).$$

Proof of Lemma B.1.1.
Given $m, l \in [\![1,n]\!]$ and an event Ω_{ml} (to be specified below) it clearly follows

$$\begin{aligned}\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\Omega_{ml}} &= \frac{\exp(-\eta n\{-\|\theta_{n,\overline{m}}\|^2_{l^2}+\mathrm{pen}^\Lambda(m)\})}{\sum_{l\in[\![1,n]\!]}\exp(-\eta n\{-\|\theta_{n,\bar{l}}\|^2_{l^2}+\mathrm{pen}^\Lambda(l)\})}\mathbb{1}_{\Omega_{ml}}\\ &\leqslant \exp\big(\eta n\{\|\theta_{n,\overline{m}}\|^2_{l^2}-\|\theta_{n,\bar{l}}\|^2_{l^2}+(\mathrm{pen}^\Lambda(l)-\mathrm{pen}^\Lambda(m))\}\big)\mathbb{1}_{\Omega_{ml}}\end{aligned} \tag{B.7}$$

We distinguish the two cases (i) $m \in [\![1,l[\![$ and (ii) $m \in]\!]l,n]\!]$. Consider first (i) $m \in [\![1,l[\![$. From (i) in Lemma 3.2.1 (with $\breve{\theta}_{\cdot} = \theta_{n,\overline{n}}$) follows that

$$\begin{aligned}\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\Omega_{ml}} &\leqslant \exp\big(\eta n\{\|\theta_{n,\overline{m}}\|^2_{l^2}-\|\theta_{n,\bar{l}}\|^2_{l^2}+(\mathrm{pen}^\Lambda(l)-\mathrm{pen}^\Lambda(m))\}\big)\mathbb{1}_{\Omega_{ml}}\\ &\leqslant \exp\big(\eta n\{\tfrac{11}{2}\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|^2_{l^2}-\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}(\mathfrak{b}_k^2(\theta^\circ)-\mathfrak{b}_l^2(\theta^\circ))+(\mathrm{pen}^\Lambda(l)-\mathrm{pen}^\Lambda(k))\}\big)\mathbb{1}_{\Omega_{kl}}\end{aligned}$$

If we define $\Omega_{ml} := \{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|^2_{l^2}<\kappa\mathfrak{R}_n^l(\theta^\circ,\Lambda)/7\}$ then the last bound together with eq. (3.4), i.e., $[\|\theta^\circ_{\underline{0}}\|^2_{l^2}+\kappa]\mathfrak{R}_n^m(\theta^\circ,\Lambda) \geqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}_m^2(\theta^\circ)\vee\mathrm{pen}^\Lambda(m)$, implies the assertion (i), that is

$$\begin{aligned}&\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\bar{l}}-\theta^\circ_{\bar{l}}\|^2_{l^2}<\kappa\mathfrak{R}_n^l(\theta^\circ,\Lambda)/7\}}\\ &\leqslant \exp\big(\eta n\{\tfrac{11}{14}\kappa\mathfrak{R}_n^l(\theta^\circ,\Lambda)+\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}_l^2(\theta^\circ)+\mathrm{pen}^\Lambda(l)-\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}_m^2(\theta^\circ)-\mathrm{pen}^\Lambda(m)\}\big)\\ &\qquad\leqslant \exp\big(\eta n\{[\tfrac{25}{14}\kappa+\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}]\mathfrak{R}_n^l(\theta^\circ,\Lambda)-\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}_m^2(\theta^\circ)-\mathrm{pen}^\Lambda(m)\}\big).\end{aligned}$$

Consider secondly (ii) $m \in]\!]l,n]\!]$. From (ii) in Lemma 3.2.1 (with $\breve{\theta}_{\cdot} = \theta_{n,\overline{n}}$) and (B.7) follows

$$\begin{aligned}\mathbb{P}_M^{(\eta)}(k)\mathbb{1}_{\Omega_{lk}} &\leqslant \exp\big(\eta n\{\|\theta_{n,\overline{m}}\|^2_{l^2}-\|\theta_{n,\bar{l}}\|^2_{l^2}+(\mathrm{pen}^\Lambda(l)-\mathrm{pen}^\Lambda(m))\}\big)\mathbb{1}_{\Omega_{ml}}\\ &\leqslant \exp\big(\eta n\{\tfrac{7}{2}\|\theta_{n,\bar{k}}-\theta^\circ_{\bar{k}}\|^2_{l^2}+\tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}(\mathfrak{b}_l^2(\theta^\circ)-\mathfrak{b}_m^2(\theta^\circ))+(\mathrm{pen}^\Lambda(l)-\mathrm{pen}^\Lambda(m))\}\big)\mathbb{1}_{\Omega_{lk}}\end{aligned}$$

If we set $\Omega_{lm} := \{\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} < \mathrm{pen}^\Lambda(m)/7\}$ then we clearly have

$$\mathbb{P}^{(\eta)}_M(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2}<\mathrm{pen}^\Lambda(m)/7\}}$$
$$\leqslant \exp\big(\eta n\{-\tfrac{1}{2}\mathrm{pen}^\Lambda(m) + \mathrm{pen}^\Lambda(l) + \tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_l(\theta^\circ) - \tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ)\}\big)$$

and hence, by exploiting $\mathfrak{b}^2_m(\theta^\circ) \geqslant 0$ and eq. (3.4) follows the assertion (ii), that is

$$\mathbb{P}^{(\eta)}_M(k)\mathbb{1}_{\{\|\theta_{n,\overline{k}}-\theta^\circ_{\overline{k}}\|^2_{l^2}<\mathrm{pen}^\Lambda(m)\}} \leqslant \exp\big(\eta n\{-\tfrac{1}{2}\mathrm{pen}^\Lambda(m) + [\tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2} + \kappa]\mathfrak{R}^l_n(\theta^\circ,\Lambda)\}\big),$$

which completes the proof. □

Proof of Lemma 3.2.3.
Consider (i). For the non trivial case $m_- > 1$ from Lemma B.1.1 (i) with $l = m^\dagger_-$ follows for all $m < m_- \leqslant m^\dagger_-$, and hence due to the definition (3.18) $\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m \geqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m_--1} > 2[\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 2\kappa]\mathfrak{R}^{m^\dagger_-}_n$. Exploiting the last bound we obtain for each $m \in [\![1, m_-[\![$

$$\mathbb{P}^{(\eta)}_M(m)\mathbb{1}_{\{\|\theta_{n,\overline{m^\dagger_-}}-\theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2}<\kappa\mathfrak{R}^{m^\dagger_-}_n/7\}}$$
$$\leqslant \exp\big(\eta n\{-\tfrac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}}{2}\,\mathfrak{b}^2_m(\theta^\circ) + [\tfrac{25\kappa}{14} + \tfrac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}}{2}]\mathfrak{R}^{m^\dagger_-}_n - \mathrm{pen}^\Lambda(m)\}\big)$$
$$\leqslant \exp\big(-\tfrac{3}{14}\eta\kappa n\mathfrak{R}^{m^\dagger_-}_n - \eta n\,\mathrm{pen}^\Lambda(m)\big)$$

which in turn with $\mathrm{pen}^\Lambda(m) = \kappa m\delta_\Lambda(m)\Lambda_+(m)n^{-1} \geqslant \kappa mn^{-1}$ and $\sum_{m\in\mathbb{N}}\exp(-\mu m) \leqslant \mu^{-1}$ for any $\mu > 0$ implies (i), that is,

$$\mathbb{P}^{(\eta)}_M([\![1, m_-[\![) \leqslant \mathbb{P}^{(\eta)}_M([\![1, m_-[\![)\mathbb{1}_{\{\|\theta_{n,\overline{m^\dagger_-}}-\theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2}<\kappa\mathfrak{R}^{m^\dagger_-}_n/7\}} + \mathbb{1}_{\{\|\theta_{n,\overline{m^\dagger_-}}-\theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2}\geqslant\kappa\mathfrak{R}^{m^\dagger_-}_n/7\}}$$
$$\leqslant \exp\big(-\tfrac{3\eta\kappa}{14}n\mathfrak{R}^{m^\dagger_-}_n\big)\sum_{k=1}^{m_--1}\exp(-\eta\kappa m) + \mathbb{1}_{\{\|\theta_{n,\overline{m^\dagger_-}}-\theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2}\geqslant\kappa\mathfrak{R}^{m^\dagger_-}_n/7\}}$$
$$\leqslant \tfrac{1}{\eta\kappa}\exp\big(-\tfrac{3\eta\kappa}{14}n\mathfrak{R}^{m^\dagger_-}_n\big) + \mathbb{1}_{\{\|\theta_{n,\overline{m^\dagger_-}}-\theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2}\geqslant\kappa\mathfrak{R}^{m^\dagger_-}_n/7\}}.$$

Consider (ii). From Lemma B.1.1 (ii) with $l = m^\dagger_+$ follows for all $m > m_+ \geqslant m^\dagger_+$, and hence due to the definition (3.18) $\mathrm{pen}^\Lambda(m) > 2[3\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 2\kappa]\mathfrak{R}^{m^\dagger_+}_n$. Thereby, we obtain for $m \in]\!]m_+, n]\!]$

$$\mathbb{P}^{(\eta)}_M(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2}<\mathrm{pen}^\Lambda(m)/7\}} \leqslant \exp\big(\eta n\{-\tfrac{1}{4}\mathrm{pen}^\Lambda(m) - \tfrac{1}{4}\mathrm{pen}^\Lambda(m) + [\tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2} + \kappa]\mathfrak{R}^{m^\dagger_+}_n\}\big)$$
$$\leqslant \exp\big(\eta n\{-\tfrac{1}{4}\mathrm{pen}^\Lambda(m)\}\big).$$

which in turn with $\operatorname{pen}^\Lambda(m) = \kappa m \delta_\Lambda(m)\Lambda_+(m)n^{-1}$ implies

$$\sum_{m\in\rrbracket m_+,n\rrbracket} \operatorname{pen}^\Lambda(m)\,\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2}\leqslant \operatorname{pen}/7\}}$$
$$\leqslant \kappa n^{-1}\sum_{m\in\rrbracket m_+,n\rrbracket} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\tfrac{\eta\kappa}{4}m\delta_\Lambda(m)\Lambda_+(m)\big) \quad \text{(B.8)}$$

Exploiting that $\sqrt{\delta_\Lambda(m)} = \frac{\log(m\Lambda_+(m)\vee(m+2))}{\log(m+2)} \geqslant 1$, $\kappa/4 \geqslant 2\log(3e)$ and $\eta \geqslant 1$, then for all $k\in\mathbb{N}$ we have $\frac{\eta\kappa}{4}k - \log(k+2) \geqslant 1$, and hence by $a\exp(-ab) \leqslant \exp(-b)$ for $a, b \geqslant 1$, it follows

$$\begin{aligned}
&\delta_\Lambda(m)m\Lambda_+(m)\exp\big(-\tfrac{\eta\kappa}{4}\delta_\Lambda(m)m\Lambda_+(m)\big)\\
&\qquad\leqslant \delta_\Lambda(m)\exp\big(-\tfrac{\eta\kappa}{4}\delta_\Lambda(m)m\Lambda_+(m)+\sqrt{\delta_\Lambda(m)}\log(m+2)\big)\\
&\qquad\leqslant \delta_\Lambda(m)\exp\big(-\delta_\Lambda(m)(\tfrac{\eta\kappa}{4}m-\log(m+2))\big)\leqslant \exp\big(-(\tfrac{\eta\kappa}{4}m-\log(m+2))\big)\\
&\qquad= (m+2)\exp\big(-\tfrac{\eta\kappa}{4}m\big).
\end{aligned}$$

Exploiting $\sum_{m\in\mathbb{N}}\mu m\exp(-\mu m)\leqslant \mu^{-1}$ and $\sum_{m\in\mathbb{N}}\mu\exp(-\mu m)\leqslant 1$ for any μ; we obtain

$$\sum_{k=m_++1}^{n}\delta_\Lambda(m)m\Lambda_+(m)\exp\big(-\tfrac{\eta\kappa}{4}\delta_\Lambda(m)m\Lambda_+(m)\big)\leqslant \sum_{k=m_++1}^{\infty}(m+2)\exp\big(-\tfrac{\eta\kappa}{4}m\big)\leqslant \tfrac{16}{\kappa^2\eta^2}+\tfrac{8}{\kappa\eta}.$$

Combining the last bound and (B.8) we obtain the assertion (ii), that is

$$\sum_{m\in\rrbracket m_+,n\rrbracket} \operatorname{pen}^\Lambda(m)\,\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2}\leqslant \operatorname{pen}/7\}}\leqslant n^{-1}\{\tfrac{16}{\kappa\eta^2}+\tfrac{8}{\eta}\}$$

which completes the proof. □

PROOF OF LEMMA 3.2.4.

By definition of $\widehat{m}$ it holds $-\|\theta_{n,\overline{\widehat{m}}}\|^2_{l^2}+\operatorname{pen}^\Lambda(\widehat{m})\leqslant -\|\theta_{n,\overline{m}}\|^2_{l^2}+\operatorname{pen}^\Lambda(m)$ for all $m\in\llbracket 1,n\rrbracket$, and hence

$$\|\theta_{n,\overline{\widehat{m}}}\|^2_{l^2}-\|\theta_{n,\overline{m}}\|^2_{l^2}\geqslant \operatorname{pen}^\Lambda(\widehat{m})-\operatorname{pen}^\Lambda(m)\text{ for all } m\in\llbracket 1,n\rrbracket. \quad \text{(B.9)}$$

Consider (i). It is sufficient to show, that $\{\widehat{m}\in\llbracket 1,m_-\llbracket\}\subseteq\{\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2}\geqslant \kappa\mathfrak{R}_n^{m_-^\dagger}/7\}$ for $m_->1$ holds. On the event $\{\widehat{m}\in\llbracket 1,m_-\llbracket\}$ holds $1\leqslant\widehat{m}<m_-\leqslant m_-^\dagger$ and thus by definition (3.18)

$$\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{\widehat{m}}(\theta^\circ)>[\|\theta^\circ_{\underline{0}}\|^2_{l^2}+4\kappa]\mathfrak{R}_n^{m_-^\dagger} \quad \text{(B.10)}$$

and due to Lemma 3.2.1 (i) also

$$\|\theta_{n,\overline{\widehat{m}}}\|^2_{l^2}-\|\theta_{n,\overline{m_-^\dagger}}\|^2_{l^2}\leqslant \tfrac{11}{2}\|\theta_{n,\overline{m_-^\dagger}}-\theta^\circ_{\overline{m_-^\dagger}}\|^2_{l^2}-\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\{\mathfrak{b}^2_{\widehat{m}}(\theta^\circ)-\mathfrak{b}^2_{m_-^\dagger}(\theta^\circ)\}. \quad \text{(B.11)}$$

Combining, (B.9) and (B.11) it follows that

$$\tfrac{11}{2}\|\theta_{n,\overline{m^\dagger_-}} - \theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2} \geqslant \mathrm{pen}^\Lambda(\widehat{m}) - \mathrm{pen}^\Lambda(m^\dagger_-) + \tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\{\mathfrak{b}^2_{\widehat{m}}(\theta^\circ) - \mathfrak{b}^2_{m^\dagger_-}(\theta^\circ)\}$$

and hence together with $\mathrm{pen}^\Lambda(\widehat{m}) \geqslant 0$, (B.10) and eq. (3.4) we obtain the claim, that is

$$\begin{aligned}\tfrac{11}{2}\|\theta_{n,\overline{m^\dagger_-}} - \theta^\circ_{\overline{m^\dagger_-}}\|^2_{l^2} &\geqslant \tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{\widehat{m}}(\theta^\circ) - \tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m^\dagger_-}(\theta^\circ) - \mathrm{pen}^\Lambda(m^\dagger_-)\\ &> [\tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 2\kappa]\mathfrak{R}_n^{m^\dagger_-} - \tfrac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m^\dagger_-}(\theta^\circ) - \mathrm{pen}^\Lambda(m^\dagger_-) \geqslant \tfrac{11}{14}\kappa\mathfrak{R}_n^{m^\dagger_-},\end{aligned}$$

and shows (i). Consider (ii). It is sufficient to show that, $\{\widehat{m} \in \rrbracket m_+, n\rrbracket\} \subseteq \{\|\theta_{n,\overline{\widehat{m}}} - \theta^\circ_{\overline{\widehat{m}}}\|^2_{l^2} \geqslant \mathrm{pen}^\Lambda(\widehat{m})/7\}$. On the event $\{\widehat{m} \in \rrbracket m_+, n\rrbracket\}$ holds $\widehat{m} > m_+ \geqslant m^\dagger_+$ and thus by definition (3.18)

$$\mathrm{pen}^\Lambda(\widehat{m}) > [6\|\theta\|^2_{l^2} + 4\kappa]\mathfrak{R}_n^{m^\dagger_+} \tag{B.12}$$

and due to Lemma 3.2.1 (ii) also

$$\|\theta_{n,\overline{\widehat{m}}}\|^2_{l^2} - \|\theta_{n,\overline{m^\dagger_+}}\|^2_{l^2} \leqslant \tfrac{7}{2}\|\theta_{n,\overline{\widehat{m}}} - \theta^\circ_{\overline{\widehat{m}}}\|^2_{l^2} + \tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\{\mathfrak{b}^2_{m^\dagger_+}(\theta^\circ) - \mathfrak{b}^2_{\widehat{m}}(\theta^\circ)\}. \tag{B.13}$$

Combining, (B.9) and (B.13) it follows that

$$\tfrac{7}{2}\|\theta_{n,\overline{\widehat{m}}} - \theta^\circ_{\overline{\widehat{m}}}\|^2_{l^2} \geqslant \mathrm{pen}^\Lambda(\widehat{m}) - \mathrm{pen}^\Lambda(m^\dagger_+) - \tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\{\mathfrak{b}^2_{m^\dagger_+}(\theta^\circ) - \mathfrak{b}^2_{\widehat{m}}(\theta^\circ)\}$$

and hence together with $\mathfrak{b}^2_{\widehat{m}}(\theta^\circ) \geqslant 0$, (B.12) and eq. (3.4) we obtain the claim, that is

$$\begin{aligned}\tfrac{7}{2}\|\theta_{n,\overline{\widehat{m}}} - \theta^\circ_{\overline{\widehat{m}}}\|^2_{l^2} &\geqslant (\tfrac{1}{2} + \tfrac{1}{2})\,\mathrm{pen}^\Lambda(\widehat{m}) - \mathrm{pen}^\Lambda(m^\dagger_+) - \tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m^\dagger_+}(\theta^\circ)\\ &> \tfrac{1}{2}\,\mathrm{pen}^\Lambda(\widehat{m}) + \tfrac{1}{2}[6\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 4\kappa]\mathfrak{R}_n^{m^\dagger_+} - \mathrm{pen}^\Lambda(m^\dagger_+) - \tfrac{3}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m^\dagger_+}(\theta^\circ) \geqslant \tfrac{1}{2}\,\mathrm{pen}^\Lambda(\widehat{m}),\end{aligned}$$

which shows (ii) and completes the proof. □

Lemma B.1.2.
Assume that Assumption 18 holds true and use the penalty described in Definition 37. Then, for all $n \in \mathbb{N}$ and $m \in \llbracket 1, n\rrbracket$ hold

(i) *let* $m_{\mathcal{C}_3} := \lfloor 3(2/\mathcal{C}_3)^2\rfloor$ *and* $n_o := 15(\mathcal{C}_5)^{-4}$ *then*
$\sum_{m=1}^n \mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} - 12\Delta_\Lambda(m)/n\right)_+ \leqslant \mathcal{C}_1 n^{-1}\left[\frac{2\mathcal{C}_2}{\mathcal{C}_3}\Lambda_+(m_{\mathcal{C}_3}) + \mathcal{C}_4 n_o\Lambda_+(n_o)\right]$

(ii) *let* $m_{\mathcal{C}_7} := \lfloor 3(2/\mathcal{C}_7)^2\rfloor$ *and* $n_o := 15(3/\mathcal{C}_8)^4$ *then*
$\sum_{m=1}^n \Delta_\Lambda(m)\,\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\Delta_\Lambda(m)/n\right) \leqslant \mathcal{C}_6\left[\Lambda_+(m_{\mathcal{C}_7})^2 m^2_{\mathcal{C}_7} + \Lambda_+(n_o)^2 n_o^2\right]$

(iii) $\mathbb{P}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\mathfrak{R}_n^m\right) \leqslant \mathcal{C}_9\left[\exp\left(\frac{-\mathcal{C}_{10} n\mathfrak{R}_n^m}{\Lambda_+(m)}\right) + (\mathcal{C}_8)^{-2}n^{-1}\right]$

Proof of Lemma B.1.2.

Consider (i). Since $\delta_\Lambda(m) \geqslant 1$ for $m \geqslant 3(\frac{2}{\mathcal{C}_3})^2$ holds $\sqrt{\delta_\Lambda(m)}m\frac{\mathcal{C}_3}{2} - \log(m+2) \geqslant 0$ and

$$\Lambda_+(m)\exp\big(-\delta_\Lambda(m)m\mathcal{C}_3\big) \leqslant \exp\big(-\delta_\Lambda(m)m\tfrac{\mathcal{C}_3}{2}\big)\exp\big(-\sqrt{\delta_\Lambda(m)}[\sqrt{\delta_\Lambda(m)}m\tfrac{\mathcal{C}_3}{2} - \log(m+2)]\big)$$
$$\leqslant \exp\big(-\delta_\Lambda(m)m\tfrac{\mathcal{C}_3}{2}\big) \leqslant \exp\big(-\tfrac{\mathcal{C}_3}{2}m\big)$$

consequently, for $m_{\mathcal{C}_3} := \lfloor 3(\frac{2}{\mathcal{C}_3})^2\rfloor$ then exploiting $\sum_{m\in\mathbb{N}}\exp(-\mu m) \leqslant \mu^{-1}$ follows

$$\sum_{m=1+m_{\mathcal{C}_3}}^{n} \Lambda_+(m)\exp\big(-\mathcal{C}_3\delta_\Lambda(m)m\big) \leqslant \sum_{m=1+m_{\mathcal{C}_3}}^{n} \exp\big(-\tfrac{\mathcal{C}_3}{2}m\big) \leqslant \tfrac{2}{\mathcal{C}_3}$$

while

$$\sum_{m=1}^{m_{\mathcal{C}_3}} \Lambda_+(m)\exp\big(-\mathcal{C}_3\delta_\Lambda(m)m\big) \leqslant \Lambda_+(m_{\mathcal{C}_3})\sum_{m=1}^{m_{\mathcal{C}_3}}\exp\big(-\mathcal{C}_3 m\big) \leqslant \frac{\Lambda_+(m_{\mathcal{C}_3})}{\mathcal{C}_3}$$

hence

$$\sum_{m=1}^{n} \Lambda_+(m)\exp\big(-\mathcal{C}_3\delta_\Lambda(m)m\big) \leqslant \tfrac{2}{\mathcal{C}_3} + \frac{\Lambda_+(m_{\mathcal{C}_3})}{\mathcal{C}_3} \leqslant \tfrac{2}{\mathcal{C}_3}\Lambda_+(m_{\mathcal{C}_3})$$

Using for all $n > n_o := 15(\mathcal{C}_5)^{-4}$ holds $\sqrt{n} \geqslant (\mathcal{C}_5)^{-1}\log(n+2)$ it follows for all $m \in [\![1, n]\!]$

$$\tfrac{m\Lambda_+(m)}{n\mathcal{C}_5}\exp\big(-\sqrt{n\delta_\Lambda(m)}\mathcal{C}_5\big) \leqslant \tfrac{1}{n}\exp\big(-\sqrt{\delta_\Lambda(m)}[\sqrt{n}\mathcal{C}_5 - \log(m+2)]\big) \leqslant \tfrac{1}{n}$$

consequently,

$$\sum_{m=1}^{n} \tfrac{m\Lambda_+(m)}{n}\exp\big(-\sqrt{n\delta_\Lambda(m)}\mathcal{C}_5\big) \leqslant \sum_{m=1}^{n}\tfrac{1}{n} \leqslant 1$$

while for $n \leqslant n_o$ with $\Lambda_+(n) \leqslant \Lambda_+(n_o)$ follows

$$\sum_{m=1}^{n} \tfrac{m\Lambda_+(m)}{n}\exp\big(-\sqrt{n\delta_\Lambda(m)}\mathcal{C}_5\big) \leqslant \Lambda_+(n)n\exp\big(-\mathcal{C}_5\sqrt{n}\big) \leqslant n_o\Lambda_+(n_o)$$

consequently, for all $n \in \mathbb{N}$ holds

$$\sum_{m=1}^{n} \tfrac{m\Lambda_+(m)}{n}\exp\big(-\mathcal{C}_5\sqrt{n\delta_\Lambda(m)}\big) \leqslant \Lambda_+(n_o)n_o$$

Combining the last two bounds and Assumption 18 (i) we obtain (i), that is

$$\sum_{m=1}^{n} \mathbb{E}\left(\|\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\|^2_{l^2} - 12\Delta_\Lambda(m)/n\right)_+$$
$$\leqslant \mathcal{C}_1\Big[\tfrac{\mathcal{C}_2}{n}\sum_{m=1}^{n}\Lambda_+(m)\exp\big(-\mathcal{C}_3\delta_\Lambda(m)m\big) + \tfrac{\mathcal{C}_4}{n}\sum_{m=1}^{n}\tfrac{m\Lambda_+(m)}{n}\exp\big(-\mathcal{C}_5\sqrt{n\delta_\Lambda(m)}\big)\Big]$$
$$\leqslant \mathcal{C}_1 n^{-1}\big[\tfrac{2}{\mathcal{C}_3}\Lambda_+(m_{\mathcal{C}_3})\|\phi\|^2_{l^1} + 4\Lambda_+(n_o)n_o\big]$$

Consider (ii). If $m \geqslant 3(2/\mathcal{C}_7)^2$ then $m \geqslant (2/\mathcal{C}_7)\log(m+2)$ and hence $m - (\mathcal{C}_7)^{-1}\log(m+2) \geqslant (\mathcal{C}_7)^{-1}\log(m+2)$ or equivalently, $\mathcal{C}_7 m - \log(m+2) \geqslant \log(m+2) \geqslant 1$ and thus

$$m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big) \leqslant \delta_\Lambda(m)\exp\big(-\delta_\Lambda(m)\,[\mathcal{C}_7 m - \log(m+2)]\big) \leqslant (m+2)\exp\big(-\mathcal{C}_7 m\big)$$

consequently, if $m > m_{\mathcal{C}_7} := \lfloor 3(2/\mathcal{C}_7)^2\rfloor$ exploiting $\sum_{m\in\mathbb{N}}(m+2)\exp(-\mu m) \leqslant \mu^{-2} + 2\mu^{-1}$ follows

$$\sum_{m=1+m_{\mathcal{C}_7}}^{n} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big) \leqslant \sum_{m=1+m_{\mathcal{C}_7}}^{n}(k+2)\exp\big(-\mathcal{C}_7 m\big) \leqslant (\mathcal{C}_7)^{-2} + 2/\mathcal{C}_7 \leqslant m_{\mathcal{C}_7}^2$$

while $\log(m\Lambda_+(m)) \leqslant \frac{1}{e}m\Lambda_+(m)$ implies $\delta_\Lambda(m) \leqslant m\Lambda_+(m)$ it follows

$$\sum_{m=1}^{m_{\mathcal{C}_7}} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big) \leqslant \delta_\Lambda(m_{\mathcal{C}_7})\Lambda_+(m_{\mathcal{C}_7})\sum_{m=1}^{m_{\mathcal{C}_7}} m\exp\big(-\mathcal{C}_7 m\big) \leqslant \delta_\Lambda(m_{\mathcal{C}_7})\Lambda_+(m_{\mathcal{C}_7})(\mathcal{C}_7)^{-2} \leqslant \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2$$

consequently for all $n \in \mathbb{N}$ we have

$$\sum_{m=1}^{n} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big) \leqslant (1 + \Lambda_+(m_{\mathcal{C}_7})^2)m_{\mathcal{C}_7}^2 \leqslant 2\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2$$

Since $\delta_\Lambda(m) \leqslant m\Lambda_+(m)$, and for all $n > n_o := \lfloor 15(3/\mathcal{C}_8)^4\rfloor$ holds $\sqrt{n} \geqslant 3/\mathcal{C}_8\log(n+2)$

$$m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big) \leqslant m^2\Lambda_+(m)^2\exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big)$$
$$\leqslant \tfrac{1}{n}\exp\big(-\sqrt{\delta_\Lambda(m)}[\sqrt{n}\mathcal{C}_8 - 2\log(m+2)] + \log(n+2)\big) \leqslant \tfrac{1}{n}\exp\big(-3\sqrt{\delta_\Lambda(m)}[\tfrac{\mathcal{C}_8\sqrt{n}}{3} - \log(n+2)]\big) \leqslant \tfrac{1}{n}$$

consequently,

$$\sum_{m=1}^{n} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big) \leqslant \sum_{m=1}^{n}\tfrac{1}{n} \leqslant 1$$

On the other hand side for $n \leqslant n_o$ with $n^b\exp(-an^{1/c}) \leqslant (\frac{cb}{ea})^{cb}$ for all $c > 0$ and $a, b \geqslant 0$ follows

$$\sum_{m=1}^{n} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big)$$
$$\leqslant n^2\delta_\Lambda(n)\Lambda_+(n)\exp\big(-\mathcal{C}_8\sqrt{n}\big) \leqslant \Lambda_+(n)^2 n^3\exp\big(-\mathcal{C}_8\sqrt{n}\big) \leqslant \Lambda_+(n_o)^2\big(3/\mathcal{C}_8\big)^6 \leqslant \Lambda_+(n_o)^2 n_o^2$$

consequently, for all $n \in \mathbb{N}$ holds

$$\sum_{m=1}^{n} m\delta_\Lambda(m)\Lambda_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big) \leqslant \Lambda_+(n_o)^2 n_o^2$$

Combining the last two bounds and Assumption 18 (ii) we obtain (ii), that is

$$\begin{aligned}\sum_{m=1}^{n} \delta_\Lambda(m)m\Lambda_+(m)\,\mathbb{P}\left(\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\Delta_\Lambda(m)/n\right)&\\ \leqslant \mathcal{C}_6\sum_{m=1}^{n}\delta_\Lambda(m)m\Lambda_+(m)\Big[\exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big)+\exp\big(-\mathcal{C}_8\sqrt{n\delta_\Lambda(m)}\big)\Big]&\\ \leqslant \mathcal{C}_6\Big[2\Lambda_+(m_{\mathcal{C}_7})^2m^2_{\mathcal{C}_7}+\Lambda_+(n_o)^2n_o^2\Big]&\end{aligned}$$

Consider (iii). Since $\frac{\mathcal{C}_{11}n\sqrt{\mathfrak{R}^m_n(\theta^\circ,\Lambda)}}{\sqrt{m\Lambda_+(m)}} \geqslant \mathcal{C}_{11}\sqrt{n\delta_\Lambda(m)} \geqslant \mathcal{C}_{11}\sqrt{n}$ and $n\exp(-\mathcal{C}_{11}\sqrt{n}) \leqslant (\mathcal{C}_{11})^2$ from Assumption 18 (iii) follows (iii), that is

$$\begin{aligned}\mathbb{P}\left(\|\theta_{n,\overline{m}}-\theta^\circ_{\overline{m}}\|^2_{l^2} \geqslant 12\mathfrak{R}^m_n(\theta^\circ,\Lambda)\right) \leqslant \mathcal{C}_9\big[\exp\big(\tfrac{-\mathcal{C}_{10}n\mathfrak{R}^m_n(\theta^\circ,\Lambda)}{\Lambda_+(m)}\big)+\exp\big(\tfrac{-\mathcal{C}_{11}n\sqrt{\mathfrak{R}^m_n(\theta^\circ,\Lambda)}}{\sqrt{m\Lambda_+(m)}}\big)\big]&\\ \leqslant \mathcal{C}_9\big[\exp\big(\tfrac{-\mathcal{C}_{10}n\mathfrak{R}^m_n(\theta^\circ,\Lambda)}{\Lambda_+(m)}\big)+(\mathcal{C}_8)^{-2}n^{-1}\big]&\end{aligned}$$

which completes the proof. □

Proof of Lemma 3.2.5

Since $\kappa/7 \geqslant 12$ and $\mathrm{pen}^\Lambda(m)/7 \geqslant 12\Delta_\Lambda(m)n^{-1}$, $m \in [\![1,n]\!]$, by exploiting Lemma B.1.2 (i), (ii) and (iii) we obtain immediately the claim (i), (ii) and (iii), respectively, which completes the proof. □

Proof of Lemma 3.2.6.

Consider firstly the aggregation using the aggregation weights $\mathbb{P}^{(\eta)}_M$ as in (3.3). Combining Lemma 3.2.5 and the upper bound given in 3.12 we obtain the existence of a constant $\mathcal{C}$ such that,

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|^2_{l^2} \leqslant \tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+)+2\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m_-}(\theta^\circ)&\\ +\mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\mathbb{1}_{\{m_->1\}}\big[\tfrac{1}{\eta}\exp\big(-18\eta n\mathfrak{R}^{m^\dagger_-}_n\big)+\exp\big(-\mathcal{C}_{10}n\mathfrak{R}^{m^\dagger_-}_n\Lambda_+(m^\dagger_-)^{-1}\big)\big]&\\ +n^{-1}\big(\eta^{-1}+\|\theta^\circ_{\underline{0}}\|^2_{l^2}\mathbb{1}_{\{m_->1\}}+\Lambda_+(m_{\mathcal{C}_3})+\Lambda_+(m_{\mathcal{C}_7})^2m^2_{\mathcal{C}_7}+\Lambda_+(n_o)^2n_o^2\big)&\quad\text{(B.14)}\end{aligned}$$

Moreover, since $1 \geqslant \Lambda_+(m^\dagger_-)^{-1}$ it holds $n\mathfrak{R}^{m^\dagger_-}_n \geqslant n\mathfrak{R}^{m^\dagger_-}_n\Lambda_+(m^\dagger_-)^{-1}$. From (B.14) with

$18\eta > \mathcal{C}_{10}$ (since $\eta \geqslant 1$ and $\mathcal{C}_{10} \leqslant 1$) follows

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+) + 2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_-}(\theta^\circ)\\ + 2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\mathcal{C}\big[\exp\big(-\mathcal{C}_{10}n\mathfrak{R}_n^{m_-^\dagger}\Lambda_+(m_-^\dagger)^{-1}\big)\big]\\ + n^{-1}\big[2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}} + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \tfrac{2}{7}\mathcal{C}_6\Lambda_+(n_o)^2\big]\end{aligned} \tag{B.15}$$

Consider secondly the aggregation using the model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.6). Combining Lemma 3.2.5 and the upper bound given in 3.13 we obtain

$$\begin{aligned}\mathbb{E}\|\theta_{n,\widehat{m}} - \theta^\circ\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+) + 2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_-}(\theta^\circ)\\ + 2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\mathcal{C}\big[\exp\big(-\mathcal{C}_{10}n\mathfrak{R}_n^{m_-^\dagger}\Lambda_+(m_-^\dagger)^{-1}\big)\big]\\ + n^{-1}\big[2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}} + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \tfrac{2}{7}\mathcal{C}_6\Lambda_+(n_o)^2\big]\end{aligned} \tag{B.16}$$

From (B.15) and (B.16) together with $n\mathfrak{R}_n^{m_-^\dagger}\Lambda_+(m_-^\dagger)^{-1} \geqslant \delta_\Lambda(m_-^\dagger)m_-^\dagger$ follows the claim, which completes the proof. □

Proof of Theorem 3.2.1.
From (3.14) follows for any $m_-^\dagger, m_+^\dagger \in [\![1,n]\!]$ and associated $m_-, m_+ \in [\![1,n]\!]$ as defined in (3.11)

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+) + 2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_-}(\theta^\circ) + \mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\big[\exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_-^\dagger)m_-^\dagger\big)\big]\\ + \mathcal{C}\big[\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}} + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\big]n^{-1}.\end{aligned} \tag{B.17}$$

We distinguish the two cases **(p)** and **(np)**. Consider first **(p)**, and hence there is $K \in \mathbb{N}_0$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ and $\mathfrak{b}_m(\theta^\circ) = 0$ for all $m \geqslant K$. Consider first $K = 0$, then $\mathfrak{b}_0(\theta^\circ) = 0$ and hence $\|\theta^\circ_{\underline{0}}\|_{l^2}^2 = 0$. From (B.17) follows

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+) + \mathcal{C}\big[\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\big]n^{-1} \tag{B.18}$$

Setting $m_+^\dagger := 1$ it follows from the definition 3.11 of m_+ that $\mathrm{pen}^\Lambda(m_+) \leqslant 4\kappa\mathfrak{R}_n^1$, where $\mathfrak{R}_n^1 = \Delta_\Lambda(1)n^{-1}$ and $\Delta_\Lambda(1) = \delta_\Lambda(1)\Lambda_+(1) \leqslant \Lambda_+(1)^2$. Thereby with numerical constant $\kappa \geqslant 84$, (B.18) implies

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}\big[\Lambda_+(1)^2 + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\big]n^{-1} \tag{B.19}$$

Consider now $K \in \mathbb{N}$, and hence $\|\theta^\circ_{\underline{0}}\|_{l^2}^2 > 0$. Let $c_{\theta^\circ} := \frac{\|\theta^\circ_{\underline{0}}\|_{l^2}^2 + 4\kappa}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)} > 1$ and $n_{\theta^\circ} := \lceil c_{\theta^\circ}\Delta_\Lambda(K)\rceil \in \mathbb{N}$. We distinguish for $n \in \mathbb{N}$ the following two cases, (a) $n \in [\![1, n_{\theta^\circ}]\!]$ and (b) $n > n_{\theta^\circ}$. Firstly, consider (a) with $n \in [\![1, n_{\theta^\circ}]\!]$, then setting $m_-^\dagger := 1$, $m_+^\dagger := 1$ we have $m_- = 1$, $1 \geqslant \mathfrak{b}_{m_-}$ and from the definition (3.18) of m_+ also $\mathrm{pen}^\Lambda(m_+) \leqslant 2[3\|\theta^\circ_{\underline{0}}\|_{l^2}^2 + 2\kappa]\mathfrak{R}_n^1 \leqslant 10\kappa\,[\|\theta^\circ_{\underline{0}}\|_{l^2}^2 \vee 1]\Lambda_+(1)^2$ exploiting $\mathfrak{b}_1 \leqslant 1 \leqslant \Delta_\Lambda(1) = \delta_\Lambda(1)\Lambda_+(1) \leqslant \Lambda_+(1)^2$.

Thereby, from B.17 follows

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2 &\leqslant \tfrac{20}{7}\kappa(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)\Lambda_+(1)^2+2\|\theta^\circ_{\underline{0}}\|_{l^2}^2+\mathcal{C}\big[\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2\big]n^{-1}\\ &\leqslant \mathcal{C}\big[(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)\Lambda_+(1)^2n+\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2\big]n^{-1}\end{aligned}$$

Moreover, for all $n\in[\![1,n_{\theta^\circ}]\!]$ with $n_{\theta^\circ}=\lceil c_{\theta^\circ}\Delta_\Lambda(K)\rceil$ and $\Delta_\Lambda(K)=K\delta_\Lambda(K)\Lambda_+(K)\leqslant K^2\Lambda_+(K)^2$ holds $n\leqslant\mathcal{C}\frac{(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)}K^2\Lambda_+(K)^2$and thereby,

$$\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2\leqslant\mathcal{C}\big[(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)\Lambda_+(1)^2\tfrac{K^2\Lambda_+(K)^2}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)}+\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2\big]n^{-1}. \tag{B.20}$$

Secondly, consider (b), i.e., $n>n_{\theta^\circ}$. Setting $m_+^\dagger:=K<\lceil c_{\theta^\circ}\Delta_\Lambda(K)\rceil=n_{\theta^\circ}$, i.e., $m_+^\dagger\in[\![1,n]\!]$, it follows $\mathfrak{b}_{m_+^\dagger}(\theta^\circ)=0$ and hence $\mathfrak{R}_n^{m_+^\dagger}=\Delta_\Lambda(K)n^{-1}$. Therefore, the definition (3.18) of m_+ implies $\mathrm{pen}^\Lambda(m_+)\leqslant[6\|\theta^\circ_{\underline{0}}\|_{l^2}^2+4\kappa]\Delta_\Lambda(K)n^{-1}\leqslant\mathcal{C}(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)K^2\Lambda_+(K)^2n^{-1}$. From (B.17) follows for all $n>n_{\theta^\circ}$ thus

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2\leqslant 2\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_-}(\theta^\circ)+\mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_-^\dagger)m_-^\dagger\big)\\ +\mathcal{C}\big[(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)K^2\Lambda_+(K)^2+\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2\big]n^{-1}.\end{aligned} \tag{B.21}$$

Since $n>n_{\theta^\circ}:=\lceil c_{\theta^\circ}\Delta_\Lambda(K)\rceil$ with $c_{\theta^\circ}=\frac{\|\theta^\circ_{\underline{0}}\|_{l^2}^2+4\kappa}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)}>1$ the defining set of $m_n^\bullet:=\max\{m\in[\![K,n]\!]:n>c_{\theta^\circ}\Delta_\Lambda(m)\}$ evenutally containing K and is hence not empty. Consequently, $m_n^\bullet\geqslant K$ and, hence $\mathfrak{b}_{m_n^\bullet}(\theta^\circ)=0$, and $\mathfrak{R}_n^{m_n^\bullet}=\Delta_\Lambda(m_n^\bullet)n^{-1}<c_{\theta^\circ}^{-1}=\frac{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2+4\kappa}$, it follows $\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)>[\|\theta^\circ_{\underline{0}}\|_{l^2}^2+4\kappa]\mathfrak{R}_n^{m_n^\bullet}$ and trivially $\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_K(\theta^\circ)=0<[\|\theta^\circ_{\underline{0}}\|_{l^2}^2+4\kappa]\mathfrak{R}_n^{m_n^\bullet}$. Therefore, setting $m_-^\dagger:=m_n^\bullet$ the definition (3.18) implies $m_-=K$ and hence $\mathfrak{b}^2_{m_-}(\theta^\circ)=\mathfrak{b}^2_K(\theta^\circ)=0$. From (B.21) follows now for all $n>n_{\theta^\circ}$ thus

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2\leqslant\mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_n^\bullet)m_n^\bullet\big)\\ +\mathcal{C}\big[(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)K^2\Lambda_+(K)^2+\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2\big]n^{-1}.\end{aligned} \tag{B.22}$$

Combining (B.20) and (B.22) for $K\geqslant 1$ with (a) $n\in[\![1,n_{\theta^\circ}[\![$ and (b) $n\geqslant n_{\theta^\circ}$, respectively, and (B.19) for $K=0$ implies for all $K\in\mathbb{N}_0$ and for all $n\in\mathbb{N}$ the claim (3.15) in case **(p)**, that is

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2\leqslant\mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\big[n^{-1}\vee\exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_n^\bullet)m_n^\bullet\big)\big]\\ +\mathcal{C}\big[\Lambda_+(1)^2\{\tfrac{(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)K^2\Lambda_+(K)^2}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)}\mathbb{1}_{K\geqslant 1}+\mathbb{1}_{K=0}\}+\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2\big]n^{-1}.\end{aligned} \tag{B.23}$$

Consider the case **(np)**. For $m_n^\dagger(\theta^\circ)\in[\![1,n]\!]$ as in 3.4 set $m_+^\dagger:=m_n^\dagger(\theta^\circ)$ and $m_-^\dagger:=m_n^\bullet\in[\![m_n^\dagger(\theta^\circ),n]\!]$ by exploiting the definition (3.11) of m_+ and m_- it follows $\mathrm{pen}^\Lambda(m_+)\leqslant 2[3\|\theta^\circ_{\underline{0}}\|_{l^2}^2+2\kappa]\mathfrak{R}_n^{m_+^\dagger}(\theta^\circ)$ and $\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_-}(\theta^\circ)\leqslant[\|\theta^\circ_{\underline{0}}\|_{l^2}^2+4\kappa]\mathfrak{R}_n^{m_n^\bullet}(\theta^\circ)$ which together with

$\mathfrak{R}_n^{m_n^\bullet}(\theta^\circ) \geqslant \mathfrak{R}_n^\dagger(\theta^\circ) = \mathfrak{R}_n^{m_+^\dagger}(\theta^\circ) \geqslant n^{-1}$ and exploiting (B.17) implies

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}(\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \vee 1)\big[\mathfrak{R}_n^{m_n^\bullet}(\theta^\circ, \Lambda) \vee \exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_n^\bullet)m_n^\bullet\big)\big] \\ + \mathcal{C}\big[\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\big]n^{-1} \quad \text{(B.24)}$$

which shows the assertion (3.16) and completes the proof. □

B.1.2 Proofs for section 3.2.1.2

Below we state the proofs of Lemma 3.2.7 and Lemma 3.2.8. The proof of Lemma 3.2.7 is based on Lemma B.1.3 given first.

Lemma B.1.3.
Consider the data-driven aggregation weights $\mathbb{P}_M^{(\eta)}$ as in (3.3). Under definition 3.4 for any $l \in [\![1, n]\!]$ with $\mathfrak{R}_n^l := \mathfrak{R}_n^l(\mathfrak{a})$ holds

(i) *for all $k \in [\![1, l[\![$ we have*
$\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\bar{l}} - \theta_{\bar{l}}^\circ\|_{l^2}^2 < \kappa\mathfrak{R}_n^l(\mathfrak{a})/7\}} \leqslant \exp\big(\eta n\{-\frac{\|\theta_{\underline{0}}^\circ\|_{l^2}^2}{2}\,\mathfrak{b}_m^2(\theta^\circ) + [\frac{25\kappa}{14} + \frac{\|\theta_{\underline{0}}^\circ\|_{l^2}^2}{2}]\mathfrak{R}_n^l - \mathrm{pen}^\Lambda(m)\}\big)$

(ii) *for all $m \in]\!]l, n]\!]$ we have*
$\mathbb{P}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,\overline{m}} - \theta_{\overline{m}}^\circ\|_{l^2}^2 < \mathrm{pen}^\Lambda(m)/7\}} \leqslant \exp\big(\eta n\{-\frac{1}{2}\,\mathrm{pen}^\Lambda(m) + [\frac{3}{2}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + \kappa]\mathfrak{R}_n^l\}\big).$

Proof of Lemma B.1.3.
The proof follows line by line the proof of Lemma B.1.1 using (3.17) rather than (3.10), and we omit the details. □

Proof of Lemma 3.2.7.
The proof follows line by line the proof of Lemma 3.2.3 using Lemma B.1.3 rather than Lemma B.1.1, and we omit the details. □

Proof of Lemma 3.2.8.
The proof follows line by line the proof of Lemma 3.2.4 using (3.17) rather than (3.10), and we omit the details. □

Proof of Theorem 3.2.2.
Keep in mind that $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \leqslant r^2$ for all $\theta^\circ \in \Theta(\mathfrak{a}, r)$. From (3.19) follows for any $\theta^\circ \in \Theta(\mathfrak{a}, r)$, $m_-^\dagger, m_+^\dagger \in [\![1, n]\!]$ and associated $m_-, m_+ \in [\![1, n]\!]$ as defined in (3.18)

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \tfrac{2}{7}\,\mathrm{pen}^\Lambda(m_+) + 2\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ) + \mathcal{C}r^2\exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_-^\dagger)m_-^\dagger\big) \\ + \mathcal{C}\big[r^2 + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\big]n^{-1} \quad \text{(B.25)}$$

For $m_n^\dagger(\mathfrak{a}) \in [\![1, n]\!]$ as in 40 set $m_+^\dagger := m_n^\dagger(\mathfrak{a})$ and $m_-^\dagger := m_n^\bullet \in [\![m_n^\dagger(\mathfrak{a}), n]\!]$ by exploiting the definition (3.18) of m_+ and m_- it follows $\mathrm{pen}^\Lambda(m_+) \leqslant 2[3r^2 + 2\kappa]\mathfrak{R}_n^{m_+^\dagger}(\mathfrak{a})$ and $\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ) \leqslant [r^2 + 4\kappa]\mathfrak{R}_n^{m_n^\bullet}(\mathfrak{a})$ which together with $\mathfrak{R}_n^{m_n^\bullet}(\mathfrak{a}) \geqslant \mathfrak{R}_n^\dagger(\mathfrak{a}) = \mathfrak{R}_n^{m_+^\dagger}(\mathfrak{a}) \geqslant n^{-1}$

and exploiting (B.25) implies

$$\sup_{\theta^\circ\in\Theta(\mathfrak{a},r)} \mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}(r^2 \vee 1) \min_{m\in[\![1,n]\!]} \left[\mathfrak{R}_n^m(\mathfrak{a}) \vee \exp\left(-\mathcal{C}_{10}\delta_\Lambda(m)m\right)\right] \\ + \mathcal{C}\left[\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_o)^2\right] n^{-1} \quad \text{(B.26)}$$

which shows the assertion (3.20) and completes the proof of Theorem 3.2.2. □

Proof of corollary 3.2.2.
Under (A) holds $\exp\left(-\mathcal{C}_{10}\delta_\Lambda(m_n^\dagger(\mathfrak{a}))m_n^\dagger(\mathfrak{a})\right) \leqslant \mathfrak{R}_n^\dagger(\mathfrak{a})$ for $n > n_{\mathfrak{a},r,\Lambda}$, while $\exp\big(-\mathcal{C}_{10}\delta_\Lambda(m_n^\dagger(\mathfrak{a}))m_n^\dagger(\mathfrak{a})\big) \leqslant 1 \leqslant n\mathfrak{R}_n^\dagger(\mathfrak{a}) \leqslant n_{\mathfrak{a},r,\Lambda}\mathfrak{R}_n^\dagger(\mathfrak{a})$ for $n \in [\![1, n_{\mathfrak{a},r,\Lambda}]\!]$. Thereby, from (3.20) with $m_n^\bullet := m_n^\dagger(\mathfrak{a})$ follows immediately the assertion $\mathcal{R}_n\big(\widehat{\theta}^{(\eta)}, \Theta(\mathfrak{a}, r), \Lambda\big) \leqslant \mathcal{C}_{\mathfrak{a},r,\Lambda}\mathfrak{R}_n^\dagger\mathfrak{a}, \Lambda$ for all $n \in \mathbb{N}$, which completes the proof of corollary 3.2.2. □

B.2 Proofs for section 3.2.2

Proof of Lemma 3.2.11
We start the proof with the observation that for all $s \in [\![1, n]\!]$ with $\mathcal{X}_s := \{|\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda\}$ and $\mathcal{X}_s^c := \{|\lambda_{n_\lambda}(s)|^2 < 1/n_\lambda\}$ holds

$$\begin{aligned}\widehat{\theta}^{(\eta)}(s) - \theta^\circ(s) &= (\lambda_{n_\lambda}^+(s)\phi_n(s) - \theta^\circ(s))\,\mathbb{P}_M^{(\eta)}([\![s, n]\!]) - \theta^\circ(s)\,\mathbb{P}_M^{(\eta)}([\![1, s[\![)\\ &= \lambda_{n_\lambda}^+(s)(\phi_n(s) - \phi(s))\,\mathbb{P}_M^{(\eta)}([\![s, n]\!])\\ &+ \lambda_{n_\lambda}^+(s)(\lambda(s) - \lambda_{n_\lambda}(s))\theta^\circ(s)\,\mathbb{P}_M^{(\eta)}([\![s, n]\!]) - \mathbb{1}_{\mathcal{X}_s}\theta^\circ(s)\,\mathbb{P}_M^{(\eta)}([\![1, s[\![) - \mathbb{1}_{\mathcal{X}_s^c}\theta^\circ(s)\end{aligned}$$

Consequently, we have

$$\begin{aligned}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 6 \sum_{s\in[\![1,n]\!]} |\lambda_{n_\lambda}^+(s)|^2|\phi_n(s) - \phi(s)|^2\,\mathbb{P}_M^{(\eta)}([\![s, n]\!])\\ + 6 \sum_{s\in[\![1,n]\!]} \mathbb{1}_{\mathcal{X}_s}|\theta^\circ(s)|^2\,\mathbb{P}_M^{(\eta)}([\![1, s[\![) + 2\sum_{s>n} |\theta^\circ(s)|^2\\ + 6 \sum_{s\in[\![1,n]\!]} |\lambda_{n_\lambda}^+(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\theta^\circ(s)|^2 + 2 \sum_{s\in[\![1,n]\!]} \mathbb{1}_{\mathcal{X}_s^c}|\theta^\circ(s)|^2. \quad \text{(B.27)}\end{aligned}$$

Consider the first r.h.s. term in (B.27). We split the sum into two parts which we bound separately. Precisely, given $\breve{\theta}_{\overline{m}} = (\mathbb{1}_{\{s\leqslant m\}}\lambda_{n_\lambda}^+(s)\phi(s))_{s\in\mathbb{Z}}$ where $\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 =$

$2\sum_{s\in\llbracket 1,m\rrbracket}|\theta_{n,n_\lambda}(s)-\breve{\theta}_{\overline{m}}(s)|^2 = 2\sum_{s\in\llbracket 1,m\rrbracket}|\lambda^+_{n_\lambda}(s)|^2|\phi_n(s)-\phi(s)|^2$ it follows

$$\begin{aligned}
2\sum_{s\in\llbracket 1,n\rrbracket} & |\lambda^+_{n_\lambda}(s)|^2(\phi_n(s)-\phi(s))^2\,\mathbb{P}^{(\eta)}_M(\llbracket s,n\rrbracket)\\
&\leqslant \|\theta_{n,n_\lambda,\overline{m_+}}-\breve{\theta}_{\overline{m_+}}\|^2_{l^2}+\sum_{l\in\rrbracket m_+,n\rrbracket}\mathbb{P}^{(\eta)}_M(l)\|\theta_{n,n_\lambda,\bar{l}}-\breve{\theta}_{\bar{l}}\|^2_{l^2}\\
&\leqslant \|\theta_{n,n_\lambda,\overline{m_+}}-\breve{\theta}_{\overline{m_+}}\|^2_{l^2}+\sum_{l\in\rrbracket m_+,n\rrbracket}\mathbb{P}^{(\eta)}_M(l)\big(\|\theta_{n,n_\lambda,\bar{l}}-\breve{\theta}_{\bar{l}}\|^2_{l^2}-\mathrm{pen}(l)/7\big)_+\\
&\quad+\tfrac{1}{7}\sum\nolimits_{l\in\rrbracket m_+,n\rrbracket}\mathbb{P}^{(\eta)}_M(l)\,\mathrm{pen}(l)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\bar{l}}-\breve{\theta}_{\bar{l}}\|^2_{l^2}\geqslant\mathrm{pen}(l)/7\}}\\
&\quad+\tfrac{1}{7}\sum\nolimits_{l\in\rrbracket m_+,n\rrbracket}\mathbb{P}^{(\eta)}_M(l)\,\mathrm{pen}(l)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\bar{l}}-\breve{\theta}_{\bar{l}}\|^2_{l^2}<\mathrm{pen}(l)/7\}}
\end{aligned}\tag{B.28}$$

Consider the second and third r.h.s. term in (B.27). Splitting the first sum into two parts we obtain

$$\begin{aligned}
2\sum_{s\in\llbracket 1,n\rrbracket} & \mathbb{1}_{\mathcal{X}_s}|\theta^\circ(s)|^2\,\mathbb{P}^{(\eta)}_M(\llbracket 1,s\llbracket)+2\sum_{s>n}|\theta^\circ(s)|^2\\
&\leqslant 2\sum_{s\in\llbracket 1,m_-\rrbracket}|\theta^\circ(s)|^2\mathbb{1}_{\mathcal{X}_s}\,\mathbb{P}^{(\eta)}_M(\llbracket 1,s\llbracket)+2\sum_{s\in\rrbracket m_-,n\rrbracket}|\theta^\circ(s)|^2+2\sum_{s>n}|\theta^\circ(s)|^2\\
&\leqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\{\mathbb{P}^{(\eta)}_M(\llbracket 1,m_-\llbracket)+\mathfrak{b}^2_{m_-}(\theta^\circ)\}
\end{aligned}\tag{B.29}$$

Combining (B.27) and the upper bounds (B.28) and (B.29) we obtain the assertion (3.21), which completes the proof. □

B.2.1 Proofs for section 3.2.2.1

Below we state the proofs of Lemma 3.2.12 and Lemma 3.2.13. The proof of Lemma 3.2.12 is based on Lemma B.2.1 given first.

Lemma B.2.1.
Consider the data-driven aggregation weights $\widehat{\mathbb{P}}^{(\eta)}_M$ as in (3.4) and the rates given in Definition 39. For any $l\in\llbracket 1,n\rrbracket$ with $\mathfrak{R}^l_n(\theta^\circ,\Lambda)=[\mathfrak{b}^2_l(\theta^\circ)\vee\Delta_\Lambda(l)\,n^{-1}]$ holds

(i) *with $\mho_l:=\Big\{1/4\leqslant\Lambda(s)^{-1}\widehat{\Lambda}(s)\leqslant 9/4,\ \forall\, s\in\llbracket 1,l\rrbracket\Big\}$ for all $k\in\llbracket 1,l\llbracket$ we have*
$$\widehat{\mathbb{P}}^{(\eta)}_M(k)\mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\bar{l}}-\breve{\theta}_{\bar{l}}\|^2_{l^2}<\widehat{\mathrm{pen}}(l)/7\right\}}\mathbb{1}_{\mho_l}\\ \leqslant \exp\big(\eta n\{[\tfrac{25}{2}\kappa+\tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|^2_{l^2}]\mathcal{R}^l_n(\theta^\circ,\Lambda)-\tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ)-\tfrac{1}{50}\,\mathrm{pen}^\Lambda(m)\}\big).$$

(ii) *with $\|\Pi^\perp_{\bar{l}}\breve{\theta}_{\overline{n}}\|^2_{l^2}=2\sum_{s=l+1}^{n}\Lambda(s)^{-1}\widehat{\Lambda}(s)|\theta^\circ(s)|^2$ for all $m\in\rrbracket l,n\rrbracket$ we have*
$$\widehat{\mathbb{P}}^{(\eta)}_M(m)\mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|^2_{l^2}<\mathrm{pen}^\Lambda(m)/7\right\}}\leqslant\exp\big(\eta n\{-\tfrac{1}{2}\,\mathrm{pen}^{\widehat{\Lambda}}(m)+\tfrac{3}{2}\|\Pi^\perp_{\bar{l}}\breve{\theta}_{\overline{n}}\|^2_{l^2}+\mathrm{pen}^{\widehat{\Lambda}}(l)\}\big).$$

Proof of Lemma B.2.1.

Given $m, l \in [\![1, n]\!]$ and an event Ω_{ml} (to be specified below) it clearly follows

$$\begin{aligned}\widehat{\mathbb{P}}^{(\eta)}_M(m)\mathbb{1}_{\Omega_{ml}} &= \frac{\exp(-\eta n\{-\|\theta_{n,n_\lambda,\overline{m}}\|^2_{l^2} + \mathrm{pen}^{\widehat{\Lambda}}(m)\})}{\sum_{l\in[\![1,n]\!]}\exp(-\eta n\{-\|\theta_{n,n_\lambda,\overline{l}}\|^2_{l^2} + \mathrm{pen}^{\widehat{\Lambda}}(l)\})}\mathbb{1}_{\Omega_{ml}} \\ &\leqslant \exp\big(\eta n\big\{\|\theta_{n,n_\lambda,\overline{m}}\|^2_{l^2} - \|\theta_{n,n_\lambda,\overline{l}}\|^2_{l^2} + (\mathrm{pen}^{\widehat{\Lambda}}(l) - \mathrm{pen}^{\widehat{\Lambda}}(m))\big\}\big)\mathbb{1}_{\Omega_{ml}} \end{aligned} \tag{B.30}$$

We distinguish the two cases (i) $m \in [\![1, l[\![$ and (ii) $m \in]\!]l, n]\!]$. Consider first (i) $m \in [\![1, l[\![$. From (i) in Lemma 3.2.1 (with $\breve{\theta}_{\overline{\bullet}} = \theta_{n,n_\lambda,\overline{n}}$ and $\theta^\circ = \breve{\theta}_{\overline{n}} = (\mathbb{1}_{\{|s|\leqslant n\}}\lambda^+_{n_\lambda}(s)\phi(s))_{s\in\mathbb{Z}})$ follows that

$$\begin{aligned}\widehat{\mathbb{P}}^{(\eta)}_M(m)\mathbb{1}_{\Omega_{ml}} &\leqslant \exp\big(\eta n\big\{\|\theta_{n,n_\lambda,\overline{m}}\|^2_{l^2} - \|\theta_{n,n_\lambda,\overline{l}}\|^2_{l^2} + (\mathrm{pen}^{\widehat{\Lambda}}(l) - \mathrm{pen}^{\widehat{\Lambda}}(m))\big\}\big)\mathbb{1}_{\Omega_{ml}} \\ &\leqslant \exp\big(\eta n\big\{\tfrac{11}{2}\|\theta_{n,n_\lambda,\overline{l}} - \breve{\theta}_{\overline{l}}\|^2_{l^2} - \tfrac{1}{2}\|\Pi_{\underline{l}\overline{m}}\breve{\theta}_{\overline{n}}\|^2_{l^2} + (\mathrm{pen}^{\widehat{\Lambda}}(l) - \mathrm{pen}^{\widehat{\Lambda}}(m))\big\}\big)\mathbb{1}_{\Omega_{kl}} \end{aligned} \tag{B.31}$$

Note that on the event $\mho_l := \big\{1/2 \leqslant |\lambda(s)\lambda^+_{n_\lambda}(s)| \leqslant 3/2,\ \forall\, s \in [\![1, l]\!]\big\}$ we have

$$\begin{aligned}&\|\Pi_{\underline{l}\overline{m}}\breve{\theta}_{\overline{n}}\|^2_{l^2}\mathbb{1}_{\mho_l} \geqslant \tfrac{1}{4}\|\Pi_{\underline{l}\overline{m}}\theta^\circ\|^2_{l^2} = \tfrac{1}{4}\|\theta^\circ_{\underline{0}}\|^2_{l^2}(\mathfrak{b}^2_m(\theta^\circ) - \mathfrak{b}^2_l(\theta^\circ)) \\ &\widehat{\Lambda}_+(l)\mathbb{1}_{\mho_l} = \max\Big\{\widehat{\Lambda}(s) = (\lambda^+_{n_\lambda}(s))^2, s \in [\![1, l]\!]\Big\}\mathbb{1}_{\mho_l} \\ &\qquad\qquad\qquad\qquad \leqslant \tfrac{9}{4}\max\left\{\Lambda(s) = \lambda(s)^{-2}, s \in [\![1, l]\!]\right\} = \tfrac{9}{4}\Lambda_+(l) \\ &\widehat{\Lambda}_+(l)\mathbb{1}_{\mho_l} \geqslant \tfrac{1}{4}\Lambda_+(l)\end{aligned}$$

Thus on $\mho_l$ holds $\frac{1}{4}l\Lambda_+(l) \vee (l+2) \leqslant l\widehat{\Lambda}_+(l) \vee (l+2) \leqslant \frac{9}{4}l\Lambda_+(l) \vee (l+2)$. Since $\sqrt{\delta_\Lambda(l)} = \frac{\log(l\Lambda_+(l)\vee(l+2))}{\log(l+2)} \geqslant 1$ for all $l \in \mathbb{N}$ hold $\frac{\log(\frac{1}{4}l\Lambda_+(l)\vee(l+2))}{\log(l+2)} \geqslant \sqrt{\delta_\Lambda(l)}\frac{\log(3/4)}{\log 3} \geqslant \frac{3}{10}\sqrt{\delta_\Lambda(l)}$ and $\frac{\log(\frac{9}{4}l\Lambda_+(l)\vee(l+2))}{\log(l+2)} \leqslant \sqrt{\delta_\Lambda(l)}\frac{\log(27/4)}{\log 3} \leqslant \frac{7}{4}\sqrt{\delta_\Lambda(l)}$ which together with $\Delta_\Lambda(l) = l\delta_\Lambda(l)\Lambda_+(l)$ and $\Delta_{\widehat{\Lambda}}(l) = l\delta_{\widehat{\Lambda}}(l)\widehat{\Lambda}_+(l)$ imply

$$\begin{aligned}&\tfrac{3}{10}\sqrt{\delta_\Lambda(l)} \leqslant \sqrt{\delta_{\widehat{\Lambda}}(l)\mathbb{1}_{\mho_l}} \leqslant \tfrac{7}{4}\sqrt{\delta_\Lambda(l)} \\ &\qquad\qquad\qquad \tfrac{1}{50}\Delta_\Lambda(l) \leqslant \tfrac{9}{400}\Delta_\Lambda(l) = l\,\tfrac{9}{100}\delta_\Lambda(l)\,\tfrac{1}{4}\Lambda_+(l) \\ &\leqslant l\,\delta_{\widehat{\Lambda}}(l)\,\widehat{\Lambda}_+(l)\mathbb{1}_{\mho_l} = \Delta_{\widehat{\Lambda}}(l)\mathbb{1}_{\mho_l} \leqslant l\,\tfrac{49}{16}\delta_\Lambda(l)\,\tfrac{9}{4}\Lambda_+(l) = \tfrac{441}{64}\Delta_\Lambda(l) \leqslant 7\Delta_\Lambda(l)\end{aligned} \tag{B.32}$$

and hence for $\mathrm{pen}^\Lambda(m) = \kappa\Delta_\Lambda(m)$ and $\mathrm{pen}^{\widehat{\Lambda}}(m) = \kappa\Delta_{\widehat{\Lambda}}(m)$ follows

$$\tfrac{1}{50}\,\mathrm{pen}^\Lambda(m) \leqslant \mathrm{pen}^{\widehat{\Lambda}}(m)\mathbb{1}_{\mho_l} \leqslant 7\,\mathrm{pen}^\Lambda(m) \quad \text{for all } m \in [\![1, l]\!] \text{ and for all } l \in \mathbb{N}. \tag{B.33}$$

If we define $\Omega_{kl} := \{\|\theta_{n,n_\lambda,\overline{l}} - \breve{\theta}_{\overline{l}}\|_{l^2}^2 < \widehat{\text{pen}^\Lambda}(l)/7\} \cap \mho_l$ then the last bounds imply

$$\begin{aligned}
&\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{l}}-\breve{\theta}_{\overline{l}}\|_{l^2}^2<\widehat{\text{pen}^\Lambda}(l)/7\}}\mathbb{1}_{\mho_l}\\
&\quad\leqslant \exp\big(\eta n\{\tfrac{11}{14}\widehat{\text{pen}^\Lambda}(l) - \tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|_{l^2}^2(\mathfrak{b}_m^2(\theta^\circ) - \mathfrak{b}_l^2(\theta^\circ)) + (\widehat{\text{pen}^\Lambda}(l) - \widehat{\text{pen}^\Lambda}(m))\}\big)\mathbb{1}_{\mho_l}\\
&\qquad\qquad = \exp\big(\eta n\{\tfrac{25}{14}\widehat{\text{pen}^\Lambda}(l) - \tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|_{l^2}^2(\mathfrak{b}_m^2(\theta^\circ) - \mathfrak{b}_l^2(\theta^\circ)) - \widehat{\text{pen}^\Lambda}(m)\}\big)\mathbb{1}_{\mho_l}\\
&\qquad\qquad\qquad \leqslant \exp\big(\eta n\{7 * \tfrac{25}{14}\text{pen}^\Lambda(l) + \tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|_{l^2}^2(\mathfrak{b}_l^2(\theta^\circ) - \mathfrak{b}_m^2(\theta^\circ)) - \tfrac{1}{50}\text{pen}^\Lambda(m)\}\big)
\end{aligned}$$

and hence, by exploiting eq. (3.4) for $\mathcal{R}_n^l(\theta^\circ, \Lambda) = [\mathfrak{b}_l^2(\theta^\circ) \vee \Delta_\Lambda(l)n^{-1}]$ follows the assertion (i), that is

$$\begin{aligned}
&\widehat{\mathbb{P}}_M^{(\eta)}(k)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{l}}-\breve{\theta}_{\overline{l}}\|_{l^2}^2<\widehat{\text{pen}^\Lambda}(l)/7\}}\mathbb{1}_{\mho_l}\\
&\qquad\qquad \leqslant \exp\big(\eta n\{[\tfrac{25}{2}\kappa + \tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|_{l^2}^2]\mathcal{R}_n^l(\theta^\circ, \Lambda) - \tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_m^2(\theta^\circ) - \tfrac{1}{50}\text{pen}^\Lambda(m)\}\big).
\end{aligned}$$

Consider secondly (ii) $m \in \rrbracket l, n\rrbracket$. From (ii) in Lemma 3.2.1 (with $\breve{\theta}_{\overline{\bullet}} = \theta_{n,n_\lambda,\overline{n}}$ and $\theta^\circ = \breve{\theta}_{\overline{n}} = (\mathbb{1}_{\{|s|\leqslant n\}}\lambda_{n_\lambda}^+(s)\phi(s))_{s\in\mathbb{Z}})$ and (B.30) follows

$$\begin{aligned}
\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\Omega_{lm}} &\leqslant \exp\big(\eta n\{\|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}^2 - \|\theta_{n,n_\lambda,\overline{l}}\|_{l^2}^2 + (\widehat{\text{pen}^\Lambda}(l) - \widehat{\text{pen}^\Lambda}(m))\}\big)\mathbb{1}_{\Omega_{lm}}\\
&\leqslant \exp\big(\eta n\{\tfrac{7}{2}\|\theta_{n,n_\lambda,\overline{k}} - \breve{\theta}_{\overline{k}}\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{k}\overline{l}}\breve{\theta}_{\overline{n}}\|_{l^2}^2 + (\widehat{\text{pen}^\Lambda}(l) - \widehat{\text{pen}^\Lambda}(m))\}\big)\mathbb{1}_{\Omega_{lk}} \qquad \text{(B.34)}
\end{aligned}$$

Keep in mind that

$$\begin{aligned}
&\|\Pi_{\underline{k}\overline{l}}\breve{\theta}_{\overline{n}}\|_{l^2}^2\mathbb{1}_{\mho_l} = 2\sum_{s=l+1}^{k}(\lambda(s)\lambda_{n_\lambda}^+(s))^2|\theta^\circ(s)|^2\\
&\qquad\qquad\qquad\qquad \leqslant 2\sum_{s=l+1}^{n}(\lambda(s)\lambda_{n_\lambda}^+(s))^2|\theta^\circ(s)|^2 = \|\Pi_{\Theta_{\overline{l}}}^{\perp}\breve{\theta}_{\overline{n}}\|_{l^2}^2.
\end{aligned}$$

If we set $\Omega_{lm} := \{\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 < \widehat{\text{pen}^\Lambda}(m)/7\}$ then we clearly have (ii), that is

$$\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|_{l^2}^2<\text{pen}^\Lambda(m)/7\}} \leqslant \exp\big(\eta n\{-\tfrac{1}{2}\widehat{\text{pen}^\Lambda}(m) + \tfrac{3}{2}\|\Pi_{\overline{l}}^{\perp}\breve{\theta}_{\overline{n}}\|_{l^2}^2 + \widehat{\text{pen}^\Lambda}(l)\}\big)$$

which completes the proof. □

PROOF OF LEMMA 3.2.12.

Consider (i). For the non trivial case $m_- > 1$ from Lemma B.2.1 (i) with $l = m_-^\dagger$ follows for all $m < m_- \leqslant m_-^\dagger$, and hence due to the definition (3.22) $\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_m^2 \geqslant \|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}_{m_--1}^2 >$

$[\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 8 * 13\kappa]\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda)$. Exploiting the last bound we obtain for each $m \in \llbracket 1, m_- \llbracket$

$$\begin{aligned}
&\widehat{\mathbb{P}}^{(\eta)}_M(m)\mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2} < \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\right\}\cap\mho_l} \\
&\qquad\qquad\leqslant \exp\left(\eta n\left\{-\tfrac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}}{8}\,\mathfrak{b}^2_m(\theta^\circ) + [\tfrac{25\kappa}{2} + \tfrac{\|\theta^\circ_{\underline{0}}\|^2_{l^2}}{8}]\mathfrak{R}_n^{m_-^\dagger} - \tfrac{1}{50}\,\mathrm{pen}^{\Lambda}(m)\right\}\right) \\
&\qquad\qquad\leqslant \exp\left(-\tfrac{1}{2}\eta\kappa n\mathfrak{R}_n^{m_-^\dagger} - \tfrac{1}{50}\eta n\,\mathrm{pen}^{\Lambda}(m)\right)
\end{aligned}$$

which in turn with $\mathrm{pen}^{\widehat{\Lambda}}(m) = \kappa m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)n^{-1} \geqslant \kappa m n^{-1}$ and $\sum_{m\in\mathbb{N}}\exp(-\mu m) \leqslant \mu^{-1}$ for any $\mu > 0$ implies (i), that is,

$$\begin{aligned}
&\mathbb{P}^{(\eta)}_M(\llbracket 1, m_- \llbracket) \\
&\leqslant \mathbb{P}^{(\eta)}_M(\llbracket 1, m_- \llbracket)\mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2} < \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\right\}\cap\mho_{m_-^\dagger}} + \mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2} \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\right\}\cup\mho^c_{m_-^\dagger}} \\
&\qquad\leqslant \exp\left(-\tfrac{\eta\kappa}{2}n\mathfrak{R}_n^{m_-^\dagger}\right)\sum_{k=1}^{m_- - 1}\exp(-\tfrac{\eta\kappa}{50}m) + \mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2} \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\right\}\cup\mho^c_{m_-^\dagger}} \\
&\qquad\qquad\leqslant \tfrac{50}{\eta\kappa}\exp\left(-\tfrac{\eta\kappa}{2}n\mathfrak{R}_n^{m_-^\dagger}\right) + \mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|^2_{l^2} \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\right\}\cup\mho^c_{m_-^\dagger}}.
\end{aligned}$$

Consider (ii). From Lemma B.2.1 (ii) with $l = m_+^\dagger$ follows for all $m > m_+ \geqslant m_+^\dagger$, and hence due to the definition (3.22) $\mathrm{pen}^{\widehat{\Lambda}}(m) > 6\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 4\,\mathrm{pen}^{\widehat{\Lambda}}(m)$. Thereby, we obtain for $m \in \rrbracket m_+, n \rrbracket$

$$\begin{aligned}
\widehat{\mathbb{P}}^{(\eta)}_M(m)\mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{l^2} < \mathrm{pen}^{\Lambda}(m)/7\right\}} &\leqslant \exp\left(\eta n\left\{-\tfrac{1}{2}\,\mathrm{pen}^{\widehat{\Lambda}}(m) + \tfrac{3}{2}\|\Pi^{\perp}_{\overline{m_+^\dagger}}\breve{\theta}_{\overline{n}}\|^2_{l^2} + \mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger)\right\}\right) \\
&\leqslant \exp\left(\eta n\left\{-\tfrac{1}{4}\,\mathrm{pen}^{\widehat{\Lambda}}(m)\right\}\right).
\end{aligned}$$

Note that $|\lambda_{n_\lambda}(s)|^2 \leqslant 1$ for all $s \in \mathbb{Z}$, hence if $|\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda$ then $\widehat{\Lambda}(s) = |\lambda^+_{n_\lambda}(s)|^2 \geqslant 1$. Thereby, $\widehat{\Lambda}(s) = |\lambda^+_{n_\lambda}(s)|^2 < 1$ implies $|\lambda_{n_\lambda}(s)|^2 < 1/n_\lambda$ and hence $\widehat{\Lambda}(s) = |\lambda^+_{n_\lambda}(s)|^2 = 0$. Thereby $1 > \widehat{\Lambda}_+(m) = \max\{|\lambda^+_{n_\lambda}(s)|^2, s \in \llbracket 1, m \rrbracket\}$ implies $\widehat{\Lambda}_+(m) = 0$, that is,

$$\{\widehat{\Lambda}_+(m) < 1\} = \{\widehat{\Lambda}_+(m) = 0\} \text{ and, } \mathrm{pen}^{\widehat{\Lambda}}(m) = \kappa\delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m) = \mathrm{pen}^{\widehat{\Lambda}}(m)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}. \tag{B.35}$$

Thereby, it follows

$$\begin{aligned}
\sum_{m\in\rrbracket m_+,n\rrbracket} &\mathrm{pen}^{\widehat{\Lambda}}(m)\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{k}}-\breve{\theta}_{\overline{k}}\|_{l^2}^2\leqslant \mathrm{pen}^{\widehat{\Lambda}}(m)/7\}}\\
&\leqslant \sum_{m\in\rrbracket m_+,n\rrbracket} \mathrm{pen}^{\widehat{\Lambda}}(m)\exp\big(-\tfrac{\eta}{4}n\,\mathrm{pen}^{\widehat{\Lambda}}(m)\big)\\
&= \sum_{m\in\rrbracket m_+,n\rrbracket} \mathrm{pen}^{\widehat{\Lambda}}(m)\exp\big(-\tfrac{\eta}{4}n\,\mathrm{pen}^{\widehat{\Lambda}}(m)\big)\{\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}+\mathbb{1}_{\{\widehat{\Lambda}_+(m)<1\}}\}\\
&= \sum_{m\in\rrbracket m_+,n\rrbracket} \mathrm{pen}^{\widehat{\Lambda}}(m)\exp\big(-\tfrac{\eta}{4}n\,\mathrm{pen}^{\widehat{\Lambda}}(m)\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}\\
&= \kappa n^{-1}\sum_{m\in\rrbracket m_+,n\rrbracket} m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\tfrac{\eta\kappa}{4}m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}} \qquad \text{(B.36)}
\end{aligned}$$

Exploiting that $\sqrt{\delta_{\widehat{\Lambda}}(m)} = \frac{\log(m\widehat{\Lambda}_+(m)\vee(m+2))}{\log(m+2)} \geqslant 1$, $\kappa/4 \geqslant 2\log(3e)$ and $\eta\geqslant 1$, then for all $k\in\mathbb{N}$ we have $\frac{\eta\kappa}{4}k-\log(k+2)\geqslant 1$, and hence by $a\exp(-ab)\leqslant\exp(-b)$ for $a,b\geqslant 1$, it follows

$$\begin{aligned}
&\delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)\exp\big(-\tfrac{\eta\kappa}{4}\delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}\\
&\quad\leqslant \delta_{\widehat{\Lambda}}(m)\exp\big(-\tfrac{\eta\kappa}{4}\delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)+\sqrt{\delta_{\widehat{\Lambda}}(m)}\log(m+2)\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}\\
&\quad\leqslant \delta_{\widehat{\Lambda}}(m)\exp\big(-\delta_{\widehat{\Lambda}}(m)(\tfrac{\eta\kappa}{4}m-\log(m+2))\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}\\
&\quad\leqslant \exp\big(-(\tfrac{\eta\kappa}{4}m-\log(m+2))\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}} = (m+2)\exp\big(-\tfrac{\eta\kappa}{4}m\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}\\
&\quad\leqslant (m+2)\exp\big(-\tfrac{\eta\kappa}{4}m\big). \qquad \text{(B.37)}
\end{aligned}$$

Exploiting $\sum_{m\in\mathbb{N}}\mu m\exp(-\mu m)\leqslant\mu^{-1}$ und $\sum_{m\in\mathbb{N}}\mu\exp(-\mu m)\leqslant 1$ we obtain

$$\begin{aligned}
\sum_{k=m_++1}^{n} &\delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)\exp\big(-\tfrac{\eta\kappa}{4}\delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)\big)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}\\
&\leqslant \sum_{k=m_++1}^{\infty}(m+2)\exp\big(-\tfrac{\eta\kappa}{4}m\big)\leqslant \tfrac{16}{\kappa^2\eta^2}+\tfrac{8}{\kappa\eta}.
\end{aligned}$$

Combining the last bound and (B.36) we obtain the assertion (ii), that is

$$\sum_{m\in\rrbracket m_+,n\rrbracket}\mathrm{pen}^{\widehat{\Lambda}}(m)\widehat{\mathbb{P}}_M^{(\eta)}(m)\mathbb{1}_{\{\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|_{l^2}^2\leqslant\mathrm{pen}^{\widehat{\Lambda}}(m)/7\}}\leqslant n^{-1}\{\tfrac{16}{\kappa\eta^2}+\tfrac{8}{\eta}\}$$

which completes the proof. □

Proof of Lemma 3.2.13.
By definition of $\widehat{m}$ it holds $-\|\theta_{n,n_\lambda,\overline{\widehat{m}}}\|_{l^2}^2+\mathrm{pen}^{\widehat{\Lambda}}(\widehat{m})\leqslant-\|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}^2+\mathrm{pen}^{\widehat{\Lambda}}(m)$ for all $m\in\llbracket 1,n\rrbracket$, and hence

$$\|\theta_{n,n_\lambda,\overline{\widehat{m}}}\|_{l^2}^2-\|\theta_{n,n_\lambda,\overline{m}}\|_{l^2}^2\geqslant\mathrm{pen}^{\widehat{\Lambda}}(\widehat{m})-\mathrm{pen}^{\widehat{\Lambda}}(m)\text{ for all }m\in\llbracket 1,n\rrbracket. \qquad \text{(B.38)}$$

Consider (i). It is sufficient to show, that $\{\widehat{m} \in [\![1, m_-[\![\} \cap \mho_{m_-^\dagger} \subseteq \{\|\theta_{n,n_\lambda,\underline{m_-^\dagger}} - \breve{\theta}_{\underline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)/7\}$ holds for $m_- > 1$. On the event $\{\widehat{m} \in [\![1, m_-[\![\}$ holds $1 \leqslant \widehat{m} < m_- \leqslant m_-^\dagger$ and thus by definition (3.22)

$$\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{\widehat{m}}^2(\theta^\circ) > [\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa]\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda) \tag{B.39}$$

and due to Lemma 3.2.1 (i) (with $\breve{\theta}_\bullet = \theta_{n,n_\lambda,\overline{n}}$ and $\theta^\circ = \breve{\theta}_{\overline{n}} = (\mathbb{1}_{\{|s|\leqslant n\}}\lambda_{n_\lambda}^+(s)\phi(s))_{s\in\mathbb{Z}})$ also

$$\|\theta_{n,n_\lambda,\overline{\widehat{m}}}\|_{l^2}^2 - \|\theta_{n,n_\lambda,\overline{m_-^\dagger}}\|_{l^2}^2 \leqslant \tfrac{11}{2}\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 - \tfrac{1}{2}\|\Pi_{\underline{m_-^\dagger}\overline{\widehat{m}}}\breve{\theta}_{\overline{n}}\|_{l^2}^2. \tag{B.40}$$

On $\{\widehat{m} \in [\![1, m_-[\![\} \cap \mho_{m_-^\dagger}$ we have

$$\|\Pi_{\underline{m_-^\dagger}\overline{\widehat{m}}}\breve{\theta}_{\overline{n}}\|_{l^2}^2\mathbb{1}_{\mho_{m_-^\dagger}} \geqslant \tfrac{1}{4}\|\Pi_{\underline{m_-^\dagger}\overline{\widehat{m}}}\theta^\circ\|_{l^2}^2 = \tfrac{1}{4}\|\theta_{\underline{0}}^\circ\|_{l^2}^2(\mathfrak{b}_{\widehat{m}}^2(\theta^\circ) - \mathfrak{b}_{m_-^\dagger}^2(\theta^\circ)) \tag{B.41}$$

Combining, (B.38), (B.40) and (B.41) it follows that

$$\tfrac{11}{2}\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(\widehat{m}) - \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger) + \tfrac{1}{8}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\{\mathfrak{b}_{\widehat{m}}^2(\theta^\circ) - \mathfrak{b}_{m_-^\dagger}^2(\theta^\circ)\}$$

and hence together with $\mathrm{pen}^{\widehat{\Lambda}}(\widehat{m}) \geqslant 0$, $\mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)\mathbb{1}_{\mho_{m_-^\dagger}} \leqslant 7\,\mathrm{pen}^{\Lambda}(m_-^\dagger)$ by (B.33), (B.39) and Definition 39 we obtain the claim, that is

$$\begin{aligned}
\tfrac{11}{2}\|\theta_{n,n_\lambda,\overline{m_-^\dagger}} - \breve{\theta}_{\overline{m_-^\dagger}}\|_{l^2}^2 &\geqslant \tfrac{1}{8}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{\widehat{m}}^2(\theta^\circ) - \tfrac{1}{8}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-^\dagger}^2(\theta^\circ) - \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)\\
&> \tfrac{1}{8}[\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa]\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda) - \tfrac{1}{8}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-^\dagger}^2(\theta^\circ) - \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger)\\
&\geqslant 13\kappa\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ, \Lambda) - \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger) \geqslant \tfrac{26}{14}\,\mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger) - \mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger) \geqslant \tfrac{11}{14}\,\mathrm{pen}^{\widehat{\Lambda}}(m_-^\dagger),
\end{aligned}$$

which shows (i). Consider (ii). It is sufficient to show that, $\{\widehat{m} \in]\!]m_+, n]\!]\} \subseteq \{\|\theta_{n,n_\lambda,\overline{\widehat{m}}} - \breve{\theta}_{\overline{\widehat{m}}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(\widehat{m})/7\}$. On the event $\{\widehat{m} \in]\!]m_+, n]\!]\}$ holds $\widehat{m} > m_+ \geqslant m_+^\dagger$ and thus by definition (3.22)

$$\mathrm{pen}^{\widehat{\Lambda}}(\widehat{m}) > 2[3\|\Pi_{\overline{m_+^\dagger}}^{\perp}\breve{\theta}_{\overline{n}}\|_{l^2}^2 + 2\,\mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger)] \tag{B.42}$$

and due to Lemma 3.2.1 (ii) (with $\breve{\theta}_\bullet = \theta_{n,n_\lambda,\overline{n}}$ and $\theta^\circ = \breve{\theta}_{\overline{n}} = (\mathbb{1}_{\{|s|\leqslant n\}}\lambda_{n_\lambda}^+(s)\phi(s))_{s\in\mathbb{Z}})$ also

$$\|\breve{\theta}_{\overline{\widehat{m}}}\|_{l^2}^2 - \|\breve{\theta}_{\overline{m_+^\dagger}}\|_{l^2}^2 \leqslant \tfrac{7}{2}\|\breve{\theta}_{\overline{\widehat{m}}} - \theta_{\overline{\widehat{m}}}^\circ\|_{l^2}^2 + \tfrac{3}{2}\|\Pi_{\underline{\widehat{m}}\overline{m_+^\dagger}}\breve{\theta}_{\overline{n}}\|_{l^2}^2. \tag{B.43}$$

Combining, (B.38) and (B.43) it follows that

$$\tfrac{7}{2}\|\theta_{n,n_\lambda,\overline{\widehat{m}}} - \breve{\theta}_{\overline{\widehat{m}}}\|_{l^2}^2 \geqslant \mathrm{pen}^{\widehat{\Lambda}}(\widehat{m}) - \mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger) - \tfrac{3}{2}\|\Pi_{\underline{\widehat{m}}\overline{m_+^\dagger}}\breve{\theta}_{\overline{n}}\|_{l^2}^2$$

and hence together with $\|\Pi_{\underline{\widehat{m}}\overline{m_+^\dagger}}\breve{\theta}_{\overline{n}}\|_{l^2}^2 \leqslant \|\Pi_{\overline{m_+^\dagger}}^{\perp}\breve{\theta}_{\overline{n}}\|_{l^2}^2$ and (B.42) we obtain the claim, that is

$$
\begin{aligned}
&\tfrac{7}{2}\|\theta_{n,n_\lambda,\widehat{m}} - \breve{\theta}_{\overline{m}}\|_{l^2}^2 \geqslant (\tfrac{1}{2}+\tfrac{1}{2})\,\mathrm{pen}^{\widehat{\Lambda}}(\widehat{m}) - \mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger) - \tfrac{3}{2}\|\Pi^\perp_{m_+^\dagger}\breve{\theta}_{\overline{n}}\|_{l^2}^2\\
&> \tfrac{1}{2}\,\mathrm{pen}^{\Lambda}(\widehat{m}) + \tfrac{1}{2}2[3\|\Pi^\perp_{m_+^\dagger}\breve{\theta}_{\overline{n}}\|_{l^2}^2 + 2\,\mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger)] - \mathrm{pen}^{\widehat{\Lambda}}(m_+^\dagger) - \tfrac{3}{2}\|\Pi^\perp_{m_+^\dagger}\breve{\theta}_{\overline{n}}\|_{l^2}^2 \geqslant \tfrac{1}{2}\,\mathrm{pen}^{\widehat{\Lambda}}(\widehat{m}),
\end{aligned}
$$

which shows (ii) and completes the proof. □

Lemma B.2.2.
Assume that Assumption 19 holds true and for $m \in \mathbb{N}$ consider $\mho_m := \{1/2 \leqslant |\lambda(s)\lambda^+_{n_\lambda}(s)| \leqslant 3/2 : \ \forall\, s \in [\![1,m]\!]\}$. Then, the following statements hold.

(i) *For all $m, n_\lambda \in \mathbb{N}$ with $\Lambda_+(k) \leqslant (4/9)n_\lambda$ holds $\mathbb{P}(\mho_m^c) \leqslant \sum_{s=1}^m \mathbb{P}\left(|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\right) \leqslant \mathcal{C}_{12} m \exp\left(-\frac{\mathcal{C}_{13} n_\lambda}{\Lambda_+(k)}\right)$.*

(ii) *Given $m \in \mathbb{N}$ let $n_\lambda(m) := \lceil 9\Lambda_+(m)/4\rceil$ then $\mathbb{P}(\mho_k^c) \leqslant 8(\mathcal{C}_{13})^{-1} m n_\lambda(m)^2 n_\lambda^{-2} \wedge 12 m n_\lambda(m) n_\lambda^{-1}$ holds true for all $n_\lambda \in \mathbb{N}$.*

(iii) *For all $m, n_\lambda \in \mathbb{N}$ with $n_\lambda \geqslant \frac{5}{\mathcal{C}_{13}} \log(m+2)\delta_\Lambda(m)\Lambda_+(m)$ holds $\mathbb{P}(\mho_m^c) \leqslant ((\frac{5}{e\mathcal{C}_{13}})^2 n_\lambda^{-2}) \wedge (\frac{2}{e\mathcal{C}_{13}} n_\lambda^{-1})$.*

Proof of Lemma B.2.2.
We start our proof with the observation that

$$
\begin{aligned}
&\{|\lambda_{n_\lambda}(s)/\lambda(s) - 1| \leqslant 1/3\} \subseteq \{||\lambda_{n_\lambda}(s)/\lambda(s)| - 1| \leqslant 1/3\}\\
&= \{1 - |\lambda_{n_\lambda}(s)/\lambda(s)| \leqslant 1/3 \text{ and } |\lambda_{n_\lambda}(s)/\lambda(s)| - 1 \leqslant 1/3\} = \{2/3 \leqslant |\lambda_{n_\lambda}(s)/\lambda(s)| \leqslant 4/3\}
\end{aligned}
$$

Moreover, if $|\lambda(s)|^2 \geqslant 9/(4n_\lambda)$, i.e., $2/3 \geqslant 1/(|\lambda(s)|\sqrt{n_\lambda})$ it follows

$$
\begin{aligned}
\{|\lambda_{n_\lambda}(s)|^2 < 1/n_\lambda\} = \{|\lambda_{n_\lambda}(s)| < 1/\sqrt{n_\lambda}\} &\subseteq \{|\lambda_{n_\lambda}(s)/\lambda(s)| < 1/(|\lambda(s)|\sqrt{n_\lambda})\}\\
&\subseteq \{|\lambda_{n_\lambda}(s)/\lambda(s)| < 2/3\} \subseteq \{|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\}
\end{aligned}
$$

Combining both inclusions, we get for $|\lambda(s)|^2 \geqslant 9/(4n_\lambda)$

$$
\begin{aligned}
\{|\lambda_{n_\lambda}(s)/\lambda(s) - 1| \leqslant 1/3\} &\subseteq \{2/3 \leqslant |\lambda_{n_\lambda}(s)/\lambda(s)| \leqslant 4/3 \text{ and } |\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda\}\\
&\subseteq \{2/3 \leqslant |\lambda_{n_\lambda}(s)/\lambda(s)| \leqslant 2 \text{ and } |\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda\}\\
&= \{1/2 \leqslant |\lambda(s)/\lambda_{n_\lambda}(s)| \leqslant 3/2 \text{ and } |\lambda_{n_\lambda}(s)|^2 \geqslant 1/n_\lambda\} = \{1/2 \leqslant |\lambda(s)\lambda^+_{n_\lambda}(s)| \leqslant 3/2\}
\end{aligned}
$$

Keeping in mind that $\Lambda(s) = |\lambda(s)|^{-2}$ and $\Lambda_+(m) = \max\{\Lambda(s), s \in [\![1,m]\!]\}$, if $\Lambda_+(m) \leqslant (4/9)n_\lambda$ then $|\lambda(s)|^2 \geqslant \Lambda_+(m)^{-1} \geqslant 9/(4n_\lambda)$ for all $s \in [\![1,m]\!]$ and hence

$$
\mho_m^c = \{1/2 \leqslant |\lambda(s)\lambda^+_{n_\lambda}(s)| \leqslant 3/2, \forall\, s \in [\![1,m]\!]\}^c \subset \bigcup_{s=1}^m \{|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\}
$$

From (iv) in Assumption 19, we obtain (i), that is

$$\mathbb{P}(\mho_m^c) \leqslant \sum_{s=1}^{m} \mathbb{P}\left(|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\right) \leqslant \mathcal{C}_{12} m \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{\Lambda_+(m)}\big).$$

Consider (ii). Given $m \in \mathbb{N}$ and $n_\lambda(m) := \lceil 9\Lambda_+(m)/4 \rceil \in \mathbb{N}$. We distinguish for $n_\lambda \in \mathbb{N}$ the cases (a) $n_\lambda \geqslant n_\lambda(m)$ and (b) $n_\lambda < n_\lambda(m)$. Consider (a). Note that $\Lambda_+(m) \leqslant (4/9)n_\lambda$ since $n_\lambda \geqslant n_\lambda(m)$, and hence (i) implies

$$\begin{aligned}\mathbb{P}(\mho_m^c) \leqslant \mathcal{C}_{12} m \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{\Lambda_+(m)}\big) &\leqslant \mathcal{C}_{12} m \underbrace{n_\lambda^2 \exp\big(-\tfrac{\mathcal{C}_{13}}{\Lambda_+(m)} n_\lambda\big)}_{\leqslant (\frac{2*(\mathcal{C}_{13})^{-1}\Lambda_+(m)}{e})^2 \text{ since } n_\lambda \geqslant 1} n_\lambda^{-2} \\ &\leqslant \mathcal{C}_{12} m \underbrace{\big(\tfrac{(\mathcal{C}_{13})^{-1}*8}{9e} * 9\Lambda_+(m)/4\big)^2}_{\leqslant 8(\mathcal{C}_{13})^{-1} n_\lambda(m)^2} n_\lambda^{-2} \leqslant 8(\mathcal{C}_{13})^{-1} m n_\lambda(m)^2 n_\lambda^{-2}.\end{aligned}$$

Analogously, we have $\mathbb{P}(\mho_m^c) \leqslant m(\frac{(\mathcal{C}_{13})^{-1}*4}{9e} 9\Lambda_+(m)/4) n_\lambda^{-1} \leqslant (6\mathcal{C}_{13})^{-1} m n_\lambda(m) n_\lambda^{-1}$, and thus $\mathbb{P}(\mho_m^c) \leqslant 8(\mathcal{C}_{13})^{-1} m n_\lambda(m)^2 n_\lambda^{-2} \wedge (6\mathcal{C}_{13})^{-1} m n_\lambda(m) n_\lambda^{-1}$ for all $n_\lambda \geqslant n_\lambda(m)$. On the other hand side consider (b) where $n_\lambda < n_\lambda(m)$ implies $\mathbb{P}(\mho_m^c) \leqslant n_\lambda(m)^2 n_\lambda^{-2} \wedge n_\lambda(m) n_\lambda^{-1}$. Combining both cases (a)-(b) for all $n_\lambda \in \mathbb{N}$ holds (ii) $\mathbb{P}(\mho_k^c) \leqslant 8(\mathcal{C}_{13})^{-1} m n_\lambda(m)^2 n_\lambda^{-2} \wedge 12 m n_\lambda(m) n_\lambda^{-1}$. Consider (iii). For all $n_\lambda, m \in \mathbb{N}$ with $n_\lambda \geqslant \frac{5}{\mathcal{C}_{13}} \log(m+2)\delta_\Lambda(m)\Lambda_+(m) \geqslant (9/4)\Lambda_+(m)$ from (i) follows

$$\begin{aligned} n_\lambda^2\, \mathbb{P}(\mho_m^c) &\leqslant m n_\lambda^2 \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{\Lambda_+(m)}\big) \\ &\leqslant \underbrace{\tfrac{n_\lambda^2}{\Lambda_+(m)^2} \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)}\big)}_{\leqslant (\frac{5}{\mathcal{C}_{13} e})^2 \text{ since } n_\lambda/\Lambda_+(m) \geqslant 1} m\Lambda_+(m)^2 \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)}\big) \\ &\leqslant (\tfrac{5}{\mathcal{C}_{13} e})^2 m\Lambda_+(m)^2 \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)}\big) \leqslant (\tfrac{5}{\mathcal{C}_{13} e})^2 \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)} + 2\delta_\Lambda(m)\log(m+2)\big) \\ &\leqslant (\tfrac{5}{\mathcal{C}_{13} e})^2 \exp\big(-\delta_\Lambda(m)\log(m+2)(\underbrace{\tfrac{\mathcal{C}_{13}}{2}\tfrac{n_\lambda}{\delta_\Lambda(m)\Lambda_+(m)\log(m+2)}}_{\geqslant \frac{\mathcal{C}_{13}}{2}\frac{5}{\mathcal{C}_{13}} > 2} - 2)\big) \leqslant (\tfrac{5}{\mathcal{C}_{13} e})^2. \end{aligned}$$

and

$$\begin{aligned} n_\lambda\, \mathbb{P}(\mho_m^c) &\leqslant m n_\lambda \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{\Lambda_+(m)}\big) \\ &\leqslant \underbrace{\tfrac{n_\lambda}{\Lambda_+(m)} \exp\big(-\tfrac{\mathcal{C}_{13}}{2}\tfrac{n_\lambda}{\Lambda_+(m)}\big)}_{\leqslant \frac{2}{e\mathcal{C}_{13}}} m\Lambda_+(m) \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)}\big) \\ &\leqslant \tfrac{2}{e\mathcal{C}_{13}} m\Lambda_+(m) \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)}\big) \leqslant \tfrac{2}{e\mathcal{C}_{13}} \exp\big(-\tfrac{\mathcal{C}_{13} n_\lambda}{2\Lambda_+(m)} + \delta_\Lambda(m)\log(m+2)\big) \\ &= \tfrac{2}{e\mathcal{C}_{13}} \exp\big(-\delta_\Lambda(m)\log(m+2)(\underbrace{\tfrac{\mathcal{C}_{13}}{2}\tfrac{n_\lambda}{\delta_\Lambda(m)\Lambda_+(m)\log(m+2)}}_{\geqslant \frac{\mathcal{C}_{13}}{2}\frac{5}{\mathcal{C}_{13}} > 2} - 1)\big) \leqslant \tfrac{2}{e\mathcal{C}_{13}}. \end{aligned}$$

which completes the proof. □

Lemma B.2.3.
Consider $\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}} = \sum_{s\in[\![-m,m]\!]} \lambda^+_{n_\lambda}(s)(\phi_n(s) - \phi(s))$. Conditionally on $\varepsilon_1, \dots, \varepsilon_{n_\lambda}$ the r.v.'s $Y_1, \dots, Y_n$ are iid. and we denote by $\mathbb{P}_{Y|\varepsilon}$ and $\mathbb{E}_{Y|\varepsilon}$ their conditional distribution and expectation, respectively. Let $\widehat{\Lambda}(s) = |\lambda^+_{n_\lambda}(s)|^2$, $\Lambda_\circ(m) = \frac{1}{m}\sum_{s\in[\![1,m]\!]} \widehat{\Lambda}(s)$, $\widehat{\Lambda}_+(m) = \max\{\widehat{\Lambda}(s), s \in [\![1,m]\!]\}$, $\kappa \geqslant 1$, $\Delta_{\widehat{\Lambda}}(m) = \delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)$ and $\sqrt{\delta_{\widehat{\Lambda}}(m)} = \frac{\log(m\widehat{\Lambda}_+(m)\vee(m+2))}{\log(m+2)} \geqslant 1$. Then there is a numerical constant $\mathcal{C}$ such that for all $n \in \mathbb{N}$ and $m \in [\![1,n]\!]$ holds

(i) *let $m_{\mathcal{C}_3} := [\lfloor 3(\frac{2}{\mathcal{C}_3})^2\rfloor \vee \mathcal{C}_2]$ and $n_{\mathcal{C}_5} := 15(\frac{1}{\mathcal{C}_5})^4$ then*

$$\sum_{m=1}^{n} \mathbb{E}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{L^2} - 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big)_+ \leqslant \mathcal{C}n^{-1}\big[(1 \vee \widehat{\Lambda}_+(m_{\mathcal{C}_3}))m_{\mathcal{C}_3} + (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_5}))\big]$$

(ii) *let $m_{\mathcal{C}_7} := \lfloor 3(\frac{2}{\mathcal{C}_7})^2\rfloor$ and $n_{\mathcal{C}_8} := \lfloor 15(3/\mathcal{C}_8)^4\rfloor$ then*

$$\sum\nolimits_{m=1}^{n} \delta_{\widehat{\Lambda}}(m)m\widehat{\Lambda}_+(m)\,\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{L^2} \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big) \leqslant \mathcal{C}_6\Big[(1 + \widehat{\Lambda}_+(m_{\mathcal{C}_7})^2)m^2_{\mathcal{C}_7} + 1 + \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2 n^2_{\mathcal{C}_8}\Big]$$

(iii)

$$\mathbb{P}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{L^2} \geqslant 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big) \leqslant 3\big[\exp\big(\tfrac{-\mathcal{C}_{11}\delta_{\widehat{\Lambda}}(m)m}{\|\phi\|_{l^1}}\big) + (\mathcal{C}_{11})^{-2}n^{-1}\big].$$

Proof of Lemma B.2.3.
Consider (i). Since $\delta_{\widehat{\Lambda}}(m) \geqslant 1$ for $m \geqslant 3(\frac{2}{\mathcal{C}_3})^2$ holds $\frac{\mathcal{C}_3\sqrt{\delta_{\widehat{\Lambda}}(m)}m}{2} - \log(m+2) \geqslant 0$ and

$$\begin{aligned}\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m\big) &\leqslant \exp\big(\tfrac{-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m}{2}\big)\exp\big(-\sqrt{\delta_{\widehat{\Lambda}}(m)}[\tfrac{\mathcal{C}_3\sqrt{\delta_{\widehat{\Lambda}}(m)}m}{2} - \log(m+2)]\big)\\ &\leqslant \exp\big(\tfrac{-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m}{2}\big) \leqslant \exp\big(-\tfrac{\mathcal{C}_3}{2}m\big)\end{aligned}$$

consequently, for $m_{\mathcal{C}_3} := [\lfloor 3(\frac{2}{\mathcal{C}_3})^2\rfloor \vee \mathcal{C}_2]$, exploiting $\sum_{m\in\mathbb{N}}\exp(-\mu m) \leqslant \mu^{-1}$ follows

$$\sum_{m=1+m_{\mathcal{C}_3}}^{n} \widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m\big) \leqslant \sum_{m=1+m_{\mathcal{C}_3}}^{n} \exp\big(-\tfrac{\mathcal{C}_3}{2}m\big) \leqslant \tfrac{2}{\mathcal{C}_3}$$

while

$$\sum_{m=1}^{m_{\mathcal{C}_3}} \widehat{\Lambda}_+(m) \exp\big(-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m\big) \leqslant \widehat{\Lambda}_+(m_{\mathcal{C}_3}) \sum_{m=1}^{m_{\mathcal{C}_3}} \exp\big(-\mathcal{C}_3 m\big) \leqslant \widehat{\Lambda}_+(m_{\mathcal{C}_3})\tfrac{1}{\mathcal{C}_3}$$

hence

$$\sum_{m=1}^{n} \widehat{\Lambda}_+(m) \exp\big(-\mathcal{C}_3\delta_{\widehat{\Lambda}}(m)m\big) \leqslant \tfrac{2}{\mathcal{C}_3} + \tfrac{\widehat{\Lambda}_+(m_\phi)}{\mathcal{C}_3} \leqslant \tfrac{1}{\mathcal{C}_3}(2+\widehat{\Lambda}_+(m_\phi))$$

Using for all $n > n_{\mathcal{C}_5} := 15(\frac{1}{\mathcal{C}_5})^4$ holds $\sqrt{n} \geqslant \frac{1}{\mathcal{C}_5}\log(n+2)$ it follows for all $m \in [\![1,n]\!]$

$$\tfrac{m\widehat{\Lambda}_+(m)}{n} \exp\big(-\mathcal{C}_5\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant \tfrac{1}{n} \exp\big(-\sqrt{\delta_{\widehat{\Lambda}}(m)}[\mathcal{C}_5\sqrt{n} - \log(m+2)]\big) \leqslant \tfrac{1}{n}$$

consequently,

$$\sum_{m=1}^{n} \tfrac{m\widehat{\Lambda}_+(m)}{n} \exp\big(-\mathcal{C}_5\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant \sum_{m=1}^{n} \tfrac{1}{n} \leqslant 1$$

while for $n \leqslant n_{\mathcal{C}_5}$ with $\widehat{\Lambda}_+(n) \leqslant \widehat{\Lambda}_+(n_{\mathcal{C}_5})$ follows

$$\sum_{m=1}^{n} \tfrac{m\widehat{\Lambda}_+(m)}{n} \exp\big(-\mathcal{C}_5\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant \widehat{\Lambda}_+(n)n \exp\big(-\mathcal{C}_5\sqrt{n}\big) \leqslant n_{\mathcal{C}_5}\widehat{\Lambda}_+(n_{\mathcal{C}_5})$$

consequently, for all $n \in \mathbb{N}$ holds

$$\sum_{m=1}^{n} \tfrac{m\widehat{\Lambda}_+(m)}{n} \exp\big(-\mathcal{C}_5\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_5})n_{\mathcal{C}_5})$$

Combining the last two bounds and Assumption 19 (i) we obtain (i) (keep in mind that $\mathcal{C}_2 \leqslant m_{\mathcal{C}_3}$ and $n_{\mathcal{C}_5} = 15(\frac{1}{\mathcal{C}_5})^4$ is a numerical constant), that is

$$\begin{aligned}\sum_{m=1}^{n} \mathbb{E}_{Y|\varepsilon}\big(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{L^2} - 12\Delta_{\widehat{\Lambda}}(m)n^{-1}\big)_+ \\ \leqslant \mathcal{C}n^{-1}\big[(1 \vee \widehat{\Lambda}_+(m_{\mathcal{C}_3}))m_{\mathcal{C}_3} + (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_5}))\big]\end{aligned}$$

Consider (ii). If $m \geqslant 3(\frac{2}{\mathcal{C}_7})^2$ then $m \geqslant (\frac{2}{\mathcal{C}_7})\log(m+2)$ and hence $m - (\mathcal{C}_7)^{-1}\log(m+2) \geqslant (\mathcal{C}_7)^{-1}\log(m+2)$ or equivalently, $\mathcal{C}_7 m - \log(m+2) \geqslant \log(m+2) \geqslant 1$ and thus similar to (B.37) it follows

$$\begin{aligned}m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m) \exp\big(-\mathcal{C}_7\delta_\Lambda(m)m\big) \leqslant \delta_\Lambda(m) \exp\big(-\delta_\Lambda(m)\,[\mathcal{C}_7 m - \log(m+2)]\big) \\ \leqslant (m+2)\exp\big(-\mathcal{C}_7 m\big)\end{aligned}$$

consequently, for $m_{\mathcal{C}_7} := \lfloor 3(\frac{2}{\mathcal{C}_7})^2 \rfloor$ exploiting $\sum_{m\in\mathbb{N}}(m+2)\exp(-\mu m) \leqslant \mu^{-2}+2\mu^{-1}$ follows

$$\sum_{m=1+m_{\mathcal{C}_7}}^{n} m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_7\delta_{\widehat{\Lambda}}(m)m\big) \leqslant \sum_{m=1+m_{\mathcal{C}_7}}^{n}(m+2)\exp\big(-\mathcal{C}_7 m\big)$$
$$\leqslant (\mathcal{C}_7)^{-2} + \tfrac{2}{\mathcal{C}_7} \leqslant m_{\mathcal{C}_7}^2$$

while $\log(m\widehat{\Lambda}_+(m))\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}} \leqslant \frac{1}{e}m\widehat{\Lambda}_+(m)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}}$ implies with (B.35) $\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m) = \delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}} \leqslant m\widehat{\Lambda}_+(m)^2\mathbb{1}_{\{\widehat{\Lambda}_+(m)\geqslant 1\}} = m\widehat{\Lambda}_+(m)^2$ it follows

$$\sum_{m=1}^{m_{\mathcal{C}_7}} m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_7\delta_{\widehat{\Lambda}}(m)m\big) \leqslant \delta_{\widehat{\Lambda}}(m_{\mathcal{C}_7})\widehat{\Lambda}_+(m_{\mathcal{C}_7})\sum_{m=1}^{m_{\mathcal{C}_7}} m\exp\big(-\mathcal{C}_7 m\big)$$
$$\leqslant \delta_{\widehat{\Lambda}}(m_{\mathcal{C}_7})\widehat{\Lambda}_+(m_{\mathcal{C}_7})(\mathcal{C}_7)^{-2} \leqslant \widehat{\Lambda}_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2$$

consequently for all $n\in\mathbb{N}$ we have

$$\sum_{m=1}^{n} m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_7\delta_{\widehat{\Lambda}}(m)m\big) \leqslant (1+\widehat{\Lambda}_+(m_{\mathcal{C}_7})^2)m_{\mathcal{C}_7}^2$$

Since $\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m) \leqslant m\widehat{\Lambda}_+(m)^2$, and for all $n > n_{\mathcal{C}_8} := \lfloor 15(3/\mathcal{C}_8)^4\rfloor$ holds $\sqrt{n} \geqslant 3/\mathcal{C}_8\log(n+2)$

$$m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant m^2\widehat{\Lambda}_+(m)^2\exp\big(-\mathcal{C}_8\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big)$$
$$\leqslant \tfrac{1}{n}\exp\big(-\sqrt{\delta_{\widehat{\Lambda}}(m)}[\mathcal{C}_8\sqrt{n}-2\log(m+2)]+\log(n+2)\big) \leqslant \tfrac{1}{n}\exp\big(-3\sqrt{\delta_{\widehat{\Lambda}}(m)}[\tfrac{\mathcal{C}_8\sqrt{n}}{3}-\log(n+2)]\big)$$
$$\leqslant \tfrac{1}{n}$$

consequently,

$$\sum_{m=1}^{n} m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant \sum_{m=1}^{n}\tfrac{1}{n} \leqslant 1$$

On the other hand side for $n \leqslant n_{\mathcal{C}_8}$ with $n^b\exp(-an^{1/c}) \leqslant (\frac{cb}{ea})^{cb}$ for all $c>0$ and $a,b\geqslant 0$ follows

$$\sum_{m=1}^{n} m\delta_{\widehat{\Lambda}}(m)\widehat{\Lambda}_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant n^2\delta_{\widehat{\Lambda}}(n)\widehat{\Lambda}_+(n)\exp\big(-\mathcal{C}_8\sqrt{n}\big)$$
$$\leqslant \widehat{\Lambda}_+(n)^2n^3\exp\big(-\mathcal{C}_8\sqrt{n}\big) \leqslant \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2\big(\tfrac{3}{\mathcal{C}_8}\big)^6 \leqslant \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2 n_{\mathcal{C}_8}^2$$

consequently, for all $n\in\mathbb{N}$ holds

$$\sum_{m=1}^{n} m\delta_{\widehat{\Lambda}}(m)\Lambda_+(m)\exp\big(-\mathcal{C}_8\sqrt{n\delta_{\widehat{\Lambda}}(m)}\big) \leqslant 1+\widehat{\Lambda}_+(n_{\mathcal{C}_8})^2n_{\mathcal{C}_8}^2$$

Combining the last two bounds and Assumption 19 (ii) we obtain (ii) (keep in mind that

$n_{\mathcal{C}_8} = 15(\mathcal{C}_8)^{-4}$ is a numerical constant), that is

$$\sum_{m=1}^{n} \delta_{\widehat{\Lambda}}(m) m \widehat{\Lambda}_+(m)\, \mathbb{P}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{L^2} \geqslant 12\Delta_{\widehat{\Lambda}}(m) n^{-1}\right) \\ \leqslant \mathcal{C}_6 \Big[(1 + \widehat{\Lambda}_+(m_{\mathcal{C}_7})^2) m^2_{\mathcal{C}_7} + 1 + \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2 n^2_{\mathcal{C}_8}\Big]$$

Consider (iii). Since $\frac{\mathcal{C}_{11} n \sqrt{\mathfrak{R}^m_n \theta^\circ, \widehat{\Lambda}}}{\sqrt{m \widehat{\Lambda}_+(m)}} \geqslant \mathcal{C}_{11}\sqrt{n \delta_{\widehat{\Lambda}}(m)} \geqslant \mathcal{C}_{11}\sqrt{n}$ and $n \exp(-\mathcal{C}_{11}\sqrt{n}) \leqslant (\mathcal{C}_{11})^{-2}$ from Assumption 19 (iii) follows (iii), that is

$$\mathbb{P}_{Y|\varepsilon}\left(\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{L^2} \geqslant 12\Delta_{\widehat{\Lambda}}(m) n^{-1}\right) \\ \leqslant 3\big[\exp\big(\tfrac{-\mathcal{C}_{11}\delta_{\widehat{\Lambda}}(m) m}{\|\phi\|_{l^1}}\big) + (\mathcal{C}_{11})^{-2} n^{-1}\big]$$

which completes the proof. □

Proof of Lemma 3.2.14.

Since $\kappa/7 \geqslant 12$ and $\mathrm{pen}^{\widehat{\Lambda}}(m)/7 \geqslant 12\Delta_{\widehat{\Lambda}}(m) n^{-1}$, $m \in [\![1, n]\!]$, by exploiting Lemma B.2.3 (i), (ii) and (iii) we obtain immediately the claim (i), (ii) and (iii), respectively, which completes the proof. □

Proof of Lemma 3.2.15.

Consider firstly the aggregation using the aggregation weights $\widehat{\mathbb{P}}^{(\eta)}_M$ as in (3.4). Combining Assumption 19 and the upper bound given in 3.23 we obtain

$$\begin{aligned} \mathbb{E}_{Y|\varepsilon}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2_{l^2} \leqslant{}& 3\,\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|^2_{l^2} + 3\|\theta^\circ_{\underline{0}}\|^2_{l^2}\, \mathfrak{b}^2_{m_-}(\theta^\circ) \\ &+ \tfrac{150}{\eta\kappa}\|\theta^\circ_{\underline{0}}\|^2_{l^2} \mathbb{1}_{\{m_- > 1\}} \exp\big(-\tfrac{3\eta\kappa}{14} n \mathfrak{R}^{m^\dagger_-}_n(\theta^\circ, \Lambda)\big) \\ &+ 3\|\theta^\circ_{\underline{0}}\|^2_{l^2} \mathbb{1}_{\{m_- > 1\}}\big[\mathcal{C}\big[\exp\big(-\mathcal{C}_{11}\delta_{\widehat{\Lambda}}(m^\dagger_-) m^\dagger_-\big)\big]\mathbb{1}_{\mho_{m^\dagger_-}} + \mathbb{1}_{\mho^c_{m^\dagger_-}}\big] \\ &+ 6\sum_{s \in [\![1,n]\!]} |\lambda^+_{n_\lambda}(s)|^2 |\lambda(s) - \lambda_{n_\lambda}(s)|^2 |\theta^\circ(s)|^2 + 2\sum_{s \in [\![1,n]\!]} \mathbb{1}_{\mathcal{X}^c_s} |\theta^\circ(s)|^2 \\ &+ \mathcal{C} n^{-1}\big[(1 \vee \widehat{\Lambda}_+(m_{\mathcal{C}_7})^2) m^2_{\mathcal{C}_7} + (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2 n^2_{\mathcal{C}_8}) + 3\|\theta^\circ_{\underline{0}}\|^2_{l^2} \mathbb{1}_{\{m_- > 1\}}\big] n_{\mathcal{C}_5} + \tfrac{16}{\kappa\eta^2} + \tfrac{8}{\eta}\big] \end{aligned} \tag{B.44}$$

Consider

$$\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|^2_{l^2} \\ = 2\sum_{s=1}^{m_+} (\lambda^+_{n_\lambda}(s))^2/n = 2\sum_{s=1}^{m_+} \widehat{\Lambda}(s)/n = 2 m_+ \Lambda_\circ(m_+)/n \leqslant 2\Delta_{\widehat{\Lambda}}(m) n^{-1},$$

where by construction $\mathrm{pen}^{\widehat{\Lambda}}(m)/7 \geqslant 12\Delta_{\widehat{\Lambda}}(m) n^{-1}$ and hence we have $\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|^2_{l^2} \leqslant \frac{1}{42}\,\mathrm{pen}^{\widehat{\Lambda}}(m_+)$. Moreover, exploiting $\max_{s \in [\![1,n]\!]} \widehat{\Lambda}(s) \leqslant n_\lambda$ and $m_+ \leqslant n$ it holds also $\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|_{l^2}^2 \leqslant 2n_\lambda$. Considering the event $\mho_{m_+^\dagger}$ and its complement $\mho_{m_+^\dagger}^c$ it follows $\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|_{l^2}^2 \leqslant 2n_\lambda \mathbb{1}_{\mho_{m_+^\dagger}^c} + \frac{1}{42}\operatorname{pen}^{\widehat{\Lambda}}(m_+)\mathbb{1}_{\mho_{m_+^\dagger}}$. Taking into account the definition (3.22) of m_+ we obtain $\mathbb{E}_{Y|\varepsilon}\|\theta_{n,n_\lambda,\overline{m_+}} - \breve{\theta}_{\overline{m_+}}\|_{l^2}^2 \leqslant 2n_\lambda \mathbb{1}_{\mho_{m_+^\dagger}^c} + \frac{1}{42}[6\|\Pi_{\overline{m_+^\dagger}}^\perp \breve{\theta}_{\overline{n}}\|_{l^2}^2 + 4\operatorname{pen}^{\widehat{\Lambda}}(m_+^\dagger)]\mathbb{1}_{\mho_{m_+^\dagger}}$ Thereby, with $\eta \geqslant 1$ and $\kappa \geqslant 1$ from (B.44) follows (keep in mind that $n_{\mathcal{C}_8}$ is a numerical constant)

$$
\begin{aligned}
\mathbb{E}_{Y|\varepsilon}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \tfrac{2}{7}\operatorname{pen}^{\widehat{\Lambda}}(m_+^\dagger)\mathbb{1}_{\mho_{m_+^\dagger}} + \tfrac{3}{7}\|\Pi_{\overline{m_+^\dagger}}^\perp \breve{\theta}_{\overline{n}}\|_{l^2}^2 \mathbb{1}_{\mho_{m_+^\dagger}} + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_-}^2(\theta^\circ)& \\
+ \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\big[\exp\big(-\tfrac{3\eta\kappa}{14} n \mathfrak{R}_n^{m_-^\dagger}(\theta^\circ,\Lambda)\big) + \exp\big(-\mathcal{C}_{11}\delta_{\widehat{\Lambda}}(m_-^\dagger)m_-^\dagger\big)\mathbb{1}_{\mho_{m_-^\dagger}}\big]& \\
+ \mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\mathbb{1}_{\mho_{m_-^\dagger}^c} + n_\lambda \mathbb{1}_{\mho_{m_+^\dagger}^c} + n^{-1}\{m_{\mathcal{C}_7}^2 n_\lambda^2 \mathbb{1}_{\mho_{m_{\mathcal{C}_7}}^c} + n_\lambda^2 \mathbb{1}_{\mho_{n_{\mathcal{C}_8}}^c}\}\big]& \\
+ 6\sum_{s\in[\![1,n]\!]} |\lambda_{n_\lambda}^+(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\theta^\circ(s)|^2 + 2\sum_{s\in[\![1,n]\!]} \mathbb{1}_{\mathcal{X}_s^c}|\theta^\circ(s)|^2& \\
+ \mathcal{C}n^{-1}\{(1 \vee \widehat{\Lambda}_+(m_{\mathcal{C}_7})^2)m_{\mathcal{C}_7}^2 \mathbb{1}_{\mho_{m_{\mathcal{C}_7}}} + (1 \vee \widehat{\Lambda}_+(n_{\mathcal{C}_8})^2)\mathbb{1}_{\mho_{n_{\mathcal{C}_8}}} + \|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\}& \quad \text{(B.45)}
\end{aligned}
$$

Employing (B.32) and (B.33) follows $\widehat{\Lambda}_+(m)\mathbb{1}_{\mho_m} \leqslant \frac{9}{4}\Lambda_+(m)$, $\delta_{\widehat{\Lambda}}(m)\mathbb{1}_{\mho_m} \geqslant \frac{9}{100}\delta_\Lambda(m)$ and $\operatorname{pen}^{\widehat{\Lambda}}(m)\mathbb{1}_{\mho_m} \leqslant 7\operatorname{pen}^{\Lambda}(m)$ for all $m \in \mathbb{N}$. Thereby, (B.45) implies

$$
\begin{aligned}
\mathbb{E}_{Y|\varepsilon}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\operatorname{pen}^{\Lambda}(m_+^\dagger) + \tfrac{3}{7}\|\Pi_{\overline{m_+^\dagger}}^\perp \breve{\theta}_{\overline{n}}\|_{l^2}^2 + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_-}^2(\theta^\circ)& \\
+ \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\big[\exp\big(-\tfrac{3\eta\kappa}{14} n \mathfrak{R}_n^{m_-^\dagger}(\theta^\circ,\Lambda)\big) + \exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100}\big)\big]& \\
+ \mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\mathbb{1}_{\mho_{m_-^\dagger}^c} + n_\lambda \mathbb{1}_{\mho_{m_+^\dagger}^c} + n^{-1}\{m_{\mathcal{C}_7}^2 n_\lambda^2 \mathbb{1}_{\mho_{m_{\mathcal{C}_7}}^c} + n_\lambda^2 \mathbb{1}_{\mho_{n_{\mathcal{C}_8}}^c}\}\big]& \\
+ 6\sum_{s\in[\![1,n]\!]} |\lambda_{n_\lambda}^+(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2|\theta^\circ(s)|^2 + 2\sum_{s\in[\![1,n]\!]} \mathbb{1}_{\mathcal{X}_s^c}|\theta^\circ(s)|^2& \\
+ \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^2 + \Lambda_+(n_{\mathcal{C}_8})^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2 \mathbb{1}_{\{m_->1\}}\}& \quad \text{(B.46)}
\end{aligned}
$$

Exploiting Lemma 1.3.1 we obtain from 9

$$
\mathbb{E}\|\Pi_{\overline{m_+^\dagger}}^\perp \breve{\theta}_{\overline{n}}\|_{l^2}^2 \leqslant 4\sum_{|s|\in]\!]m_+,n]\!]} |\theta^\circ(s)|^2 \leqslant 4\|\theta_{\underline{0}}^\circ\|_{l^2}^2\, \mathfrak{b}_{m_+^\dagger}^2(\theta^\circ)
$$

from (ii) and $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) := \sum_{s\in\mathbb{N}} \theta^\circ(s)^2[1 \wedge \Lambda(s)/n_\lambda]$ as defined in (1.6)

$$
\sum_{s\in[\![1,n]\!]} |\theta^\circ(s)|^2\, \mathbb{E}\,|\lambda_{n_\lambda}^+(s)|^2|\lambda(s) - \lambda_{n_\lambda}(s)|^2 \leqslant 4\mathcal{C}_4 \mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)
$$

and from (i)

$$
\sum_{s\in[\![1,n]\!]} \mathbb{P}(\mathcal{X}_s^c)|\theta^\circ(s)|^2 \leqslant 4\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)
$$

The last bounds imply together with

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger)+\tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_+^\dagger}^2(\theta^\circ)+3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ +\mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\big[\exp\big(-\tfrac{3\eta\kappa}{14}n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ,\Lambda)\big)+\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100}\big)\big]\\ +\mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho_{m_-^\dagger}^c)+n_\lambda\,\mathbb{P}(\mho_{m_+^\dagger}^c)+n^{-1}\{m_{\mathcal{C}_7}^2n_\lambda^2\,\mathbb{P}(\mho_{m_{\mathcal{C}_7}}^c)+n_\lambda^2\,\mathbb{P}(\mho_{n_o}^c)\}\big]\\ +24\mathcal{C}_4\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)+8\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)\\ +\mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_o)^2+\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}\end{aligned}\tag{B.47}$$

Moreover, since $n\mathfrak{R}_n^{m_-^\dagger}(\theta^\circ,\Lambda)\geqslant\delta_\Lambda(m_-^\dagger)m_-^\dagger$. From (B.47) with $\frac{3\eta\kappa}{14}>\frac{9\mathcal{C}_{11}}{100}$, follows,

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger)+\tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_+^\dagger}^2(\theta^\circ)+3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ +\mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100\|\phi\|_{l^1}}\big)\\ +\mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho_{m_-^\dagger}^c)+n_\lambda\,\mathbb{P}(\mho_{m_+^\dagger}^c)+n^{-1}\{m_{\mathcal{C}_7}^2n_\lambda^2\,\mathbb{P}(\mho_{m_{\mathcal{C}_7}}^c)+n_\lambda^2\,\mathbb{P}(\mho_{n_{\mathcal{C}_8}}^c)\}\big]\\ +\mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)+\mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^2+\Lambda_+(n_{\mathcal{C}_8})^2+\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}\end{aligned}\tag{B.48}$$

Exploiting Lemma B.2.2 (ii) there is a numerical constant $\mathcal{C}$ such that for all $n_\lambda,m\in\mathbb{N}$ holds $\mathbb{P}(\mho_m^c)\leqslant\mathcal{C}m\Lambda_+(m)^2n_\lambda^{-2}$ and hence, $n_\lambda^2\,\mathbb{P}(\mho_{m_{\mathcal{C}_7}}^c)\leqslant\mathcal{C}m_{\mathcal{C}_7}\Lambda_+(m_{\mathcal{C}_7})^2$ and $n_\lambda^2\,\mathbb{P}(\mho_{n_{\mathcal{C}_8}}^c)\leqslant\mathcal{C}n_{\mathcal{C}_8}\Lambda_+(n_{\mathcal{C}_8})^2$, thereby from (B.47) follows the assertion (3.25), that is, (keep in mind that $n_{\mathcal{C}_8}$ is a numerical constant)

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger)+\tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_+^\dagger}^2(\theta^\circ)+3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ +\mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100}\big)+\mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho_{m_-^\dagger}^c)+n_\lambda\,\mathbb{P}(\mho_{m_+^\dagger}^c)\big]\\ +\mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)+\mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3+\Lambda_+(n_{\mathcal{C}_8})^2+\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}\end{aligned}\tag{B.49}$$

Consider secondly the aggregation using the model selection weights $\mathbb{P}_M^{(\infty)}$ as in (3.8). Combining Lemma 3.2.14 and the upper bound given in 3.24 we obtain

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)}-\theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger)+\tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_+^\dagger}^2(\theta^\circ)+3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ +\mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100}\big)+\mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho_{m_-^\dagger}^c)+n_\lambda\,\mathbb{P}(\mho_{m_+^\dagger}^c)\big]\\ +\mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda)+\mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3+\Lambda_+(n_{\mathcal{C}_8})^2+\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}\end{aligned}\tag{B.50}$$

From (B.49) and (B.50) together with $n\mathfrak{R}_n^{m_-^\dagger}\Lambda_+(m_-^\dagger)^{-1}\geqslant\delta_\Lambda(m_-^\dagger)m_-^\dagger$ follows the claim (3.14), which completes the proof. □

Proof of Theorem 3.2.3.

From (3.25) follows for any $m_-^\dagger,m_+^\dagger\in[\![1,n]\!]$ and associated $m_-,m_+\in[\![1,n]\!]$ as defined in

(3.22)

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger) + \tfrac{12}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_+^\dagger}^2(\theta^\circ) + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ + \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100\|\phi\|_{l^1}}\big) + \mathcal{C}\big[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho_{m_-^\dagger}^c) + n_\lambda\,\mathbb{P}(\mho_{m_+^\dagger}^c)\big]\\ + \mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}\end{aligned}\tag{B.51}$$

We destinguish the two cases (p) and (np). Consider first (p), and hence there is $K \in \mathbb{N}_0$ with $1 \geqslant \mathfrak{b}_{[K-1]}(\theta^\circ) > 0$ and $\mathfrak{b}_m(\theta^\circ) = 0$ for all $m \geqslant K$. Consider first $K = 0$, then $\mathfrak{b}_0(\theta^\circ) = 0$ and hence $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 = 0$ and $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) = 0$. From (B.51) follows

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\,\mathrm{pen}^\Lambda(m_+^\dagger) + \mathcal{C}n_\lambda\,\mathbb{P}(\mho_{m_+^\dagger}^c) + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}\tag{B.52}$$

Setting $m_+^\dagger := 1$ it follows $\mathrm{pen}^\Lambda(m_+^\dagger) = \kappa\Delta_\Lambda(1)n^{-1} = \kappa\delta_\Lambda(1)\Lambda_+(1)n^{-1} \leqslant \kappa\Lambda_+(1)^2n^{-1}$ and exploiting Lemma B.2.2 (ii) there is a numerical constant $\mathcal{C}$ such that for all $n_\lambda \in \mathbb{N}$ holds $\mathbb{P}(\mho_{m_+^\dagger}^c) \leqslant \mathcal{C}\Lambda_+(1)^2n_\lambda^{-2}$. Thereby with numerical constant $\kappa \geqslant 84$, (B.52) implies for all $n, n_\lambda \in \mathbb{N}$

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}\Lambda_+(1)^2n_\lambda^{-1} + \mathcal{C}n^{-1}\{\Lambda_+(1)^2 + \Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}\tag{B.53}$$

Consider now $K \in \mathbb{N}$, and hence $\|\theta_{\underline{0}}^\circ\|_{l^2}^2 > 0$. Let $c_{\theta^\circ} := \frac{\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa}{\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ)} > 1$ and $n_{\theta^\circ} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor \in \mathbb{N}$. We distinguish for $n \in \mathbb{N}$ the following two cases, (a) $n \in [\![1, n_{\theta^\circ}]\!]$ and (b) $n > n_{\theta^\circ}$. Firstly, consider (a) with $n \in [\![1, n_{\theta^\circ}]\!]$, then setting $m_-^\dagger := 1$, $m_+^\dagger := 1$ we have $m_- = 1$, $1 \geqslant \mathfrak{b}_{m_-}$ and $1 \leqslant \Delta_\Lambda(1) = \delta_\Lambda(1)\Lambda_+(1) \leqslant \Lambda_+(1)^2$. Thereby, from (B.51) follows

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\kappa\Lambda_+(1)^2n^{-1} + \tfrac{33}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + \mathcal{C}\big[n_\lambda\,\mathbb{P}(\mho_1^c)\big]\\ + \mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}\end{aligned}$$

Exploiting Lemma B.2.2 (ii) there is a numerical constant $\mathcal{C}$ such that for all $n_\lambda \in \mathbb{N}$ holds $\mathbb{P}(\mho_1^c) \leqslant \mathcal{C}\Lambda_+(1)^2n_\lambda^{-2}$, which together with $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) \leqslant \|\theta_{\underline{0}}^\circ\|_{l^2}^2\Lambda_+(K)n_\lambda^{-1}$ implies

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\kappa\Lambda_+(1)^2n^{-1} + \tfrac{33}{7}\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + \mathcal{C}\big[\Lambda_+(1)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2\Lambda_+(K)\big]n_\lambda^{-1}\\ + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}\end{aligned}$$

Moreover, for all $n \in [\![1, n_{\theta^\circ}]\!]$ with $n_{\theta^\circ} = \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor$ and $\Delta_\Lambda(K) = K\delta_\Lambda(K)\Lambda_+(K) \leqslant K^2\Lambda_+(K)^2$ holds $n \leqslant \mathcal{C}\frac{(\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \vee 1)}{\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ)}K^2\Lambda_+(K)^2$ and thereby, for all $n \in [\![1, n_{\theta^\circ}]\!]$ and for all $n_\lambda \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}\big[(\|\theta_{\underline{0}}^\circ\|_{l^2}^2 \vee 1)\Lambda_+(1)^2\tfrac{K^2\Lambda_+(K)^2}{\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ)} + \Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\big]n^{-1}\\ + \mathcal{C}\big[\Lambda_+(1)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2\Lambda_+(K)\big]n_\lambda^{-1}.\end{aligned}\tag{B.54}$$

Secondly, consider (b), i.e., $n > n_{\theta^\circ}$. Setting $m_+^\dagger := K < \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor = n_{\theta^\circ}$, i.e., $m_+^\dagger \in [\![1,n]\!]$, it follows $\mathfrak{b}_{m_+^\dagger}(\theta^\circ) = 0$ and $\mathrm{pen}[m_+^\dagger] = \kappa\Delta_\Lambda(K)n^{-1} \leqslant \kappa K^2\Lambda_+(K)^2n^{-1}$. From (B.51) follows for all $n > n_{\theta^\circ}$ thus

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 &\leqslant 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ)\\ &\quad + \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\left(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100\|\phi\|_{l^1}}\right) + \mathcal{C}\left[\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho_{m_-^\dagger}^c) + n_\lambda\,\mathbb{P}(\mho_K^c)\right]\\ &+ \mathcal{C}\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) + \mathcal{C}n^{-1}\{K^2\Lambda_+(K)^2n^{-1} + \Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}.\end{aligned}$$

Exploiting Lemma B.2.2 (ii) there is a numerical constant $\mathcal{C}$ such that for all $n_\lambda \in \mathbb{N}$ holds $\mathbb{P}(\mho_K^c) \leqslant \mathcal{C}K\Lambda_+(K)^2n_\lambda^{-2}$, which together with $\mathcal{R}_{n_\lambda}^\dagger(\theta^\circ,\Lambda) \leqslant \|\theta_{\underline{0}}^\circ\|_{l^2}^2\Lambda_+(K)n_\lambda^{-1}$ implies

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 &\leqslant \mathcal{C}n^{-1}\{K^2\Lambda_+(K)^2n^{-1} + \Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\}\\ &\quad + 3\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{m_-}^2(\theta^\circ) + \mathcal{C}\|\theta_{\underline{0}}^\circ\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\{\exp\left(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100\|\phi\|_{l^1}}\right) + \mathbb{P}(\mho_{m_-^\dagger}^c)\}\\ &\quad + \mathcal{C}n_\lambda^{-1}\{K\Lambda_+(K)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2\Lambda_+(K)\} \end{aligned}\tag{B.55}$$

In order to control the terms involving $m_-^\dagger$ and m_- we destinguish for $n_\lambda \in \mathbb{N}$ with $n_\lambda(\theta^\circ,\Lambda) := \lfloor 289\log(K+2)\delta_\Lambda(K)\Lambda_+(K)\rfloor$ the following two cases cases, (b-i) $n_\lambda \in [\![1, n_\lambda(\theta^\circ,\Lambda)]\!]$ and (b-ii) $n_\lambda > n_\lambda(\theta^\circ,\Lambda)$. Consider first (b-i) $n_\lambda \in [\![1, n_\lambda(\theta^\circ,\Lambda)]\!]$. We set $m_-^\dagger = 1$ and hence $m_- = 1$. Thereby, with $\mathfrak{b}_1^2(\theta^\circ) \leqslant 1$, $\log(K+2) \leqslant \frac{K+2}{e} \leqslant 2K$, $\delta_\Lambda(m)\Lambda_+(m) \leqslant K\Lambda_+(K)^2$, and hence $n_\lambda(\theta^\circ,\Lambda) \leqslant \mathcal{C}K^2\Lambda_+(K)^2$, from (B.55) follows for all $n_\lambda \in [\![1, n_\lambda(\theta^\circ,\Lambda)[\![$

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 &\leqslant \mathcal{C}n^{-1}\{K^2\Lambda_+(K)^2n^{-1} + \Lambda_+(m_{\mathcal{C}_7})^2m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}\\ &\quad + \mathcal{C}n_\lambda^{-1}\{K\Lambda_+(K)^2 + \|\theta_{\underline{0}}^\circ\|_{l^2}^2(K^2\Lambda_+(K)^2 + \Lambda_+(K))\}\end{aligned}\tag{B.56}$$

Consider now (b-ii) $n_\lambda > n_\lambda(\theta^\circ,\Lambda)$. Note that for all $n_\lambda > n_\lambda(K,\Lambda)$ the defining set of $m_{n_\lambda}^\bullet := \max\{m \in [\![K, n_\lambda]\!] : 289\log(m+2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$ is not empty, where obviously for each $m_-^\dagger \in [\![K, m_{n_\lambda}^\bullet]\!]$ holds $n_\lambda \geqslant 289\log(m_-^\dagger+2)\delta_\Lambda(m_-^\dagger)\Lambda_+(m_-^\dagger)$, and thus from Lemma B.2.2 (iii) follows $\mathbb{P}(\mho_{m_-^\dagger}^c) \leqslant 53n_\lambda^{-1}$. Since also $n > n_{\theta^\circ} := \lfloor c_{\theta^\circ}\Delta_\Lambda(K)\rfloor$ with $c_{\theta^\circ} := \frac{\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa}{\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ)} > 1$ the defining set of $m_n^\bullet := \max\{m \in [\![K,n]\!] : n > c_{\theta^\circ}\Delta_\Lambda(m)\}$ is not empty. Consequently, for all $m_-^\dagger \in [\![K, m_n^\bullet]\!]$ holds $m_-^\dagger \geqslant K$ and, hence $\mathfrak{b}_{m_-^\dagger}(\theta^\circ) = 0$, and $\mathfrak{R}_n^{m_-^\dagger} = \Delta_\Lambda(m_-^\dagger)n^{-1} < c_{\theta^\circ}^{-1} = \frac{\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ)}{\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa}$, it follows $\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_{[K-1]}^2(\theta^\circ) > [\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa]\mathfrak{R}_n^{m_-^\dagger}$ and trivially $\|\theta_{\underline{0}}^\circ\|_{l^2}^2\,\mathfrak{b}_K^2(\theta^\circ) = 0 < [\|\theta_{\underline{0}}^\circ\|_{l^2}^2 + 104\kappa]\mathfrak{R}_n^{m_-^\dagger}$. Therefore, the definition (3.22) implies $m_- = K$ and hence $\mathfrak{b}_{m_-}^2(\theta^\circ) = \mathfrak{b}_K^2(\theta^\circ) = 0$. Selecting $m_-^\dagger := m_n^\bullet \wedge m_{n_\lambda}^\bullet$ we have $\mathbb{P}(\mho_{m_-^\dagger}^c) \leqslant 53n_\lambda^{-1}$, $m_- = K$ and $\mathfrak{b}_{m_-}^2(\theta^\circ) = 0$, such that from (B.55) follows for all

$n_\lambda > n_\lambda(\theta^\circ, \Lambda)$ and $n > n_{\theta^\circ,\Lambda}$

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}n^{-1}\{K^2\Lambda_+(K)^2n^{-1} + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2 + \|\theta^\circ_{\underline{0}}\|_{l^2}^2\}$$
$$+ \mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\{\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_n^\bullet\wedge m_{n_\lambda}^\bullet)m_n^\bullet\wedge m_{n_\lambda}^\bullet}{100\|\phi\|_{l^1}}\big)\} + \mathcal{C}n_\lambda^{-1}\{K\Lambda_+(K)^2 + \|\theta^\circ_{\underline{0}}\|_{l^2}^2\Lambda_+(K)\} \quad \text{(B.57)}$$

Combining (B.56) and (B.57) for the cases (b-i) $n_\lambda \in [\![1, n_\lambda(\theta^\circ,\Lambda)]\!]$ and (b-ii) $n_\lambda > n_\lambda(\theta^\circ, \Lambda)$ we obtain for all $n > n_{\theta^\circ,\Lambda}$ and for all $n_\lambda \in \mathbb{N}$

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\big[n^{-1} \vee n_\lambda^{-1} \vee \exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_n^\bullet\wedge m_{n_\lambda}^\bullet)m_n^\bullet\wedge m_{n_\lambda}^\bullet}{100\|\phi\|_{l^1}}\big)\big]$$
$$+ \mathcal{C}n^{-1}\{K^2\Lambda_+(K)^2 + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\} + \mathcal{C}n_\lambda^{-1}(1 \vee \|\theta^\circ_{\underline{0}}\|_{l^2}^2)K\Lambda_+(K)^2 \quad \text{(B.58)}$$

Combining (B.54) and (B.58) for $K \in \mathbb{N}$ with (a) $n \in [\![1, n_{\theta^\circ,\Lambda}]\!]$ and (b) $n > n_{\theta^\circ,\Lambda}$, respectively, and (B.53) for $K = 0$, we obtain for all $K \in \mathbb{N}_0$ and for all $n, n_\lambda \in \mathbb{N}$

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\big[n^{-1} \vee n_\lambda^{-1} \vee \exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_n^\bullet\wedge m_{n_\lambda}^\bullet)m_n^\bullet\wedge m_{n_\lambda}^\bullet}{100\|\phi\|_{l^1}}\big)\big]$$
$$+ \mathcal{C}n^{-1}\{\Lambda_+(1)^2\{\tfrac{(\|\theta^\circ_{\underline{0}}\|_{l^2}^2\vee 1)K^2\Lambda_+(K)^2}{\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{[K-1]}(\theta^\circ)}\mathbb{1}_{\{K\geqslant 1\}} + \mathbb{1}_{\{K=0\}}\} + \Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}$$
$$+ \mathcal{C}n_\lambda^{-1}\{(1 \vee \|\theta^\circ_{\underline{0}}\|_{l^2}^2)K\Lambda_+(K)^2\mathbb{1}_{\{K\geqslant 1\}} + \Lambda_+(1)^2\mathbb{1}_{\{K=0\}}\}. \quad \text{(B.59)}$$

Consider the case **(np)**. We destinguish for $n_\lambda \in \mathbb{N}$ with $n_\lambda(\Lambda) := \lfloor 289\log(3)\delta_\Lambda(1)\Lambda_+(1)\rfloor$ the following two cases, (a) $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$ and (b) $n_\lambda > n_\lambda(\Lambda)$. Consider firstly the case (a) $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$. We set $m_+^\dagger = m_-^\dagger = 1$, and hence $m_- = 1$, $\mathfrak{b}_1^2(\theta^\circ) \leqslant 1$, $\mathrm{pen}^\Lambda(1) \leqslant \kappa\Lambda_+(1)^2 n^{-1}$, $\Lambda_+(1)^2 \leqslant \Lambda_+(n_o)^2$, $n_\lambda(\Lambda) \leqslant \mathcal{C}\Lambda_+(1)^2$ and due to Lemma B.2.2 (ii) $\mathbb{P}(\mho_1^c) \leqslant \mathcal{C}\Lambda_+(1)^2 n_\lambda^{-2}$. Thereby, (B.51) implies for all $n \in \mathbb{N}$ and $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant \mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}(1\vee\|\theta^\circ_{\underline{0}}\|_{l^2}^2)\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2\}n^{-1} \quad \text{(B.60)}$$

Consider secondly (b) $n_\lambda > n_\lambda(\Lambda)$. We set $m_{n_\lambda}^\bullet := \max\{m \in [\![1, n_\lambda]\!] : 289\log(m + 2)\delta_\Lambda(m)\Lambda_+(m) \leqslant n_\lambda\}$, where the defining set containing 1 is not empty. For each $m \in [\![1, m_{n_\lambda}^\bullet]\!]$ holds $n_\lambda \geqslant 289\log(m + 2)\delta_\Lambda(m)\Lambda_+(m)$, and thus from Lemma B.2.2 (iii) follows $\mathbb{P}(\mho_m^c) \leqslant 11226 n_\lambda^{-2}$. For $m_n^\dagger \in [\![1, n]\!]$ as in eq. (3.4) let $m_+^\dagger := m_n^\dagger \wedge m_{n_\lambda}^\bullet$, where $\mathrm{pen}^\Lambda(m_n^\dagger \wedge m_{n_\lambda}^\bullet) \leqslant \mathrm{pen}^\Lambda(m_n^\dagger) \leqslant \mathfrak{R}_n^{m_n^\dagger}(\theta^\circ,\Lambda)$, then from (B.51) follows

$$\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|_{l^2}^2 \leqslant 2\mathfrak{R}_n^{m_n^\dagger}(\theta^\circ,\Lambda) + \tfrac{12}{7}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_n^\dagger\wedge m_{n_\lambda}^\bullet}(\theta^\circ) + 3\|\theta^\circ_{\underline{0}}\|_{l^2}^2\,\mathfrak{b}^2_{m_-}(\theta^\circ)$$
$$+ \mathcal{C}\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\exp\big(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_-^\dagger)m_-^\dagger}{100\|\phi\|_{l^1}}\big) + \mathcal{C}\big[\|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\,\mathbb{P}(\mho^c_{m_-^\dagger}) + n_\lambda^{-1}\big]$$
$$+ \mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2 m_{\mathcal{C}_7}^3 + \Lambda_+(n_o)^2 + \|\theta^\circ_{\underline{0}}\|_{l^2}^2\mathbb{1}_{\{m_->1\}}\} \quad \text{(B.61)}$$

Let $m_n^\bullet := \arg\min\{\mathfrak{R}_n^m(\theta^\circ,\Lambda) \vee \exp\big(\tfrac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_7}}\big) : m \in [\![1, n]\!]\}$, where $m_n^\bullet \in [\![m_n^\dagger, 1]\!]$ by definition of $m_n^\dagger$. Setting $m_-^\dagger := m_n^\bullet \wedge m_{n_\lambda}^\bullet$ from Lemma B.2.2 (iii) follows $\mathbb{P}(\mho^c_{m_-^\dagger}) \leqslant 53 n_\lambda^{-1}$, while m_- as in definition (3.22) satisfies $\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_{m_-}(\theta^\circ) \leqslant [\|\theta^\circ_{\underline{0}}\|^2_{l^2} + 104\kappa]\mathcal{R}_n^{m_n^\bullet \wedge m_{n_\lambda}^\bullet}(\theta^\circ, \Lambda)$, where

$$\mathcal{R}_n^{m_n^\bullet \wedge m_{n_\lambda}^\bullet}(\theta^\circ,\Lambda) \leqslant \mathcal{R}_n^{m_n^\bullet}(\theta^\circ,\Lambda) + \mathfrak{b}^2_{m_n^\dagger \wedge m_{n_\lambda}^\bullet}(\theta^\circ);$$
$$\mathcal{R}_n^{m_n^\dagger}(\theta^\circ,\Lambda) \leqslant \mathcal{R}_n^{m_n^\bullet}(\theta^\circ,\Lambda)$$
$$\text{and } \mathfrak{b}^2_{m_n^\bullet \wedge m_{n_\lambda}^\bullet}(\theta^\circ) \leqslant \mathfrak{b}^2_{m_n^\dagger \wedge m_{n_\lambda}^\bullet}(\theta^\circ).$$

Thereby, we obtain for all $n \in \mathbb{N}$ and $n_\lambda > n_\lambda(\Lambda)$

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2_{l^2} \leqslant{}& \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\{\mathfrak{R}_n^{m_n^\bullet}(\theta^\circ,\Lambda) + \mathfrak{b}^2_{m_n^\dagger \wedge m_{n_\lambda}^\bullet}(\theta^\circ)\} \\ &+ \mathcal{C}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\mathbb{1}_{\{m_->1\}}\exp\left(\tfrac{-9\mathcal{C}_{11}\delta_\Lambda(m_n^\bullet \wedge m_{n_\lambda}^\bullet)m_n^\bullet \wedge m_{n_\lambda}^\bullet}{100\|\phi\|_{l^1}}\right) + \mathcal{C}\left[\|\theta^\circ_{\underline{0}}\|^2_{l^2}\mathbb{1}_{\{m_->1\}}n_\lambda^{-1} + n_\lambda^{-1}\right] \\ &+ \mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2 m^3_{\mathcal{C}_7} + \Lambda_+(n_o)^2 + \|\theta^\circ_{\underline{0}}\|^2_{l^2}\mathbb{1}_{\{m_->1\}}\}\end{aligned} \tag{B.62}$$

Since $\mathfrak{R}_n^{m_n^\bullet}(\theta^\circ,\Lambda) \geqslant n^{-1}$ and $\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) \geqslant \frac{1}{2}\|\theta^\circ_{\underline{0}}\|^2_{l^2} n_\lambda^{-1}$ it follows for all $n \in \mathbb{N}$ and $n_\lambda > n_\lambda(\Lambda)$

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2_{l^2} \leqslant{}& \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2}) \min_{m \in [\![1,n]\!]}\{\mathfrak{R}_n^m(\theta^\circ,\Lambda) \vee \exp\left(\tfrac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_7}}\right)\} \\ &+ \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\{\mathfrak{b}^2_{m_n^\dagger \wedge m_{n_\lambda}^\bullet}(\theta^\circ) \vee \exp\left(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\mathcal{C}_7}}\right)\} \\ &+ \mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}n_\lambda^{-1} + \mathcal{C}n^{-1}\{\Lambda_+(m_{\mathcal{C}_7})^2 m^3_{\mathcal{C}_7} + \Lambda_+(n_o)^2\}\end{aligned} \tag{B.63}$$

Combining (B.60) and (B.63) for the cases (a) $n_\lambda \in [\![1, n_\lambda(\Lambda)]\!]$ and (b) $n_\lambda > n_\lambda(\Lambda)$ we obtain for all $n, n_\lambda \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2_{l^2} \leqslant{}& \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2}) \min_{m \in [\![1,n]\!]}\{\mathfrak{R}_n^m(\theta^\circ,\Lambda) \vee \exp\left(\tfrac{-\delta_\Lambda(m)m}{m_{\mathcal{C}_7}}\right)\}\mathbb{1}_{\{n_\lambda > n_\lambda(\Lambda)\}} \\ &+ \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\{\mathfrak{b}^2_{m_n^\dagger \wedge m_{n_\lambda}^\bullet}(\theta^\circ) \vee \exp\left(\tfrac{-\delta_\Lambda(m_{n_\lambda}^\bullet)m_{n_\lambda}^\bullet}{m_{\mathcal{C}_7}}\right)\}\mathbb{1}_{\{n_\lambda > n_\lambda(\Lambda)\}} \\ &+ \mathcal{C}\mathcal{R}^\dagger_{n_\lambda}(\theta^\circ,\Lambda) + \mathcal{C}(1 \vee \|\theta^\circ_{\underline{0}}\|^2_{l^2})\Lambda_+(1)^2 n_\lambda^{-1} + \mathcal{C}\{\Lambda_+(m_{\mathcal{C}_7})^2 m^3_{\mathcal{C}_7} + \Lambda_+(n_o)^2\}n^{-1}\end{aligned} \tag{B.64}$$

which shows the assertion (3.27) and completes the proof of Theorem 3.2.3. □

B.2.2 Proofs for section 3.2.2.2

Below we state the proofs of Lemma 3.2.16. The proof of Lemma 3.2.16 is based on Lemma B.2.4 given first.

Lemma B.2.4.
Consider the data-driven aggregation weights $\widehat{\mathbb{P}}^{(\eta)}_M$ as in (3.4) and the maximal rates Definition 40. For any $l \in [\![1,n]\!]$ with $\mathfrak{R}^l_n(\mathfrak{a},\Lambda) = [\mathfrak{a}(l) \vee \Delta_\Lambda(l)\, n^{-1}]$ holds

(i) *with $\mho_l := \left\{1/4 \leqslant \Lambda(s)^{-1}\widehat{\Lambda}(s) \leqslant 9/4,\ \forall\, s \in [\![1,l]\!]\right\}$ for all $k \in [\![1,l[\![$ we have*

$\widehat{\mathbb{P}}^{(\eta)}_M(k)\mathbb{1}_{\left\{\|\theta_{n,n_\lambda,\bar{l}} - \breve{\theta}_{\bar{l}}\|^2_{l^2} < \mathrm{pen}^{\widehat{\Lambda}}(l)/7\right\}}\mathbb{1}_{\mho_l}$

$$\leqslant \exp\big(\eta n\big\{[\tfrac{25}{2}\kappa + \tfrac{1}{8}r^2]\mathcal{R}_n^l(\mathfrak{a},\Lambda) - \tfrac{1}{8}\|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ) - \tfrac{1}{50}\operatorname{pen}^\Lambda(m)\big\}\big).$$

(ii) *with* $\|\Pi_{\overline{l}}^{\perp}\breve{\theta}_{\overline{n}}\|^2_{l^2} = 2\sum_{s=l+1}^{n}\Lambda(s)^{-1}\widehat{\Lambda}(s)|\theta^\circ(s)|^2$ *for all* $m \in \rrbracket l, n\rrbracket$ *we have*
$$\widehat{\mathbb{P}}^{(\eta)}_M(m)\mathbb{1}_{\big\{\|\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\|^2_{l^2}<\operatorname{pen}^\Lambda(m)/7\big\}} \leqslant \exp\big(\eta n\big\{-\tfrac{1}{2}\operatorname{pen}^{\widehat{\Lambda}}(m)+\tfrac{3}{2}\|\Pi_{\overline{l}}^{\perp}\breve{\theta}_{\overline{n}}\|^2_{l^2}+\operatorname{pen}^{\widehat{\Lambda}}(l)\big\}\big).$$

Proof of Lemma B.2.4.
The assertion (i) follows from Lemma B.2.1 (i) using that $r^2\mathfrak{R}_n^m(\mathfrak{a},\Lambda) \geqslant \|\theta^\circ_{\underline{0}}\|^2_{l^2}\,\mathfrak{b}^2_m(\theta^\circ)$ uniformely for all $\theta^\circ \in \Theta(\mathfrak{a}, r)$ and for all $m \in \mathbb{N}$. The assertion (ii) equals Lemma B.2.1 (ii). □

Proof of Lemma 3.2.16.
The proof follows line by line the proof of Lemma 3.2.12 using Lemma B.2.4 rather than Lemma B.2.1, and we omit the details. □

Appendix C

Proofs for section 3.4

The next assertion provides our key arguments in order to control the deviations of the reminder terms. Both inequalities are due to Talagrand (1996), the formulation of the first part can be found for example in Klein and Rio (2005), while the second part is based on equation (5.13) in Corollary 2 in Birgé and Massart (1995) and stated in this form for example in Comte and Merlevède (2002).

Lemma C.0.1.
(Talagrand's inequalities) Let $Z_1, \dots, Z_n$ be independent $\mathcal{Z}$-valued random variables and let $\overline{\nu_{[x]}} = n^{-1}\sum_{i=1}^n [\nu_{[x]}(Z_i) - \mathbb{E}(\nu_{[x]}(Z_i))]$ for $\nu_{[x]}$ belonging to a countable class $\{\nu_{[x]}, [x] \in \mathcal{H}\}$ of measurable functions. Then,

$$\mathbb{E}\big(\sup_{[x]\in\mathcal{H}} |\overline{\nu_{[x]}}|^2 - 6H^2\big)_+ \leqslant C[\frac{v}{n}\exp(\frac{-nH^2}{6v}) + \frac{h^2}{n^2}\exp(\frac{-KnH}{h})] \tag{C.1}$$

$$\mathbb{P}\big(\sup_{[x]\in\mathcal{H}} |\overline{\nu_{[x]}}| \geqslant 2H + \psi\big) \leqslant 3\exp\big[-Kn(\frac{\psi^2}{v} \wedge \frac{\psi}{h})\big] \leqslant 3\big[\exp\big(\frac{-Kn\psi^2}{v}\big) + \exp\big(\frac{-Kn\psi}{h}\big)\big] \tag{C.2}$$

for any $\psi > 0$, with numerical constants $K = (\sqrt{2}-1)/(21\sqrt{2})$ and $C > 0$ and where

$$\sup_{[x]\in\mathcal{H}}\sup_{z\in\mathcal{Z}} |\nu_{[x]}(z)| \leqslant h, \qquad \mathbb{E}\big(\sup_{[x]\in\mathcal{H}} |\overline{\nu_{[x]}}|\big) \leqslant H, \qquad \sup_{[x]\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_{[x]}(Z_i)) \leqslant v.$$

□

Remark C.0.1 *Keeping the bounds (C.1) and (C.2) in mind, let us specify particular choices for the constants ψ and K. We choose $\psi = \sqrt{2}(\sqrt{3}-\sqrt{2})H = \frac{(\sqrt{6}-\sqrt{4})(\sqrt{6}+\sqrt{4})}{(\sqrt{6}+\sqrt{4})}H = \frac{\sqrt{2}}{(\sqrt{3}+\sqrt{2})}H$, and hence $\sqrt{2}\sqrt{3}H = \sqrt{2}\sqrt{2}H + \sqrt{2}(\sqrt{3}-\sqrt{2})H$. Moreover, we have $K\frac{2}{(\sqrt{3}+\sqrt{2})^2} = \frac{(\sqrt{2}-1)}{(21\sqrt{2})}\frac{2}{(\sqrt{3}+\sqrt{2})^2} = \frac{(2-\sqrt{2})}{21(\sqrt{3}+\sqrt{2})^2} \geqslant \frac{1}{400}$ and $K\frac{\sqrt{2}}{(\sqrt{3}+\sqrt{2})} = \frac{\sqrt{2}-1}{21(\sqrt{3}+\sqrt{2})} \geqslant \frac{1}{200}$ and $K \geqslant \frac{1}{100}$. The* *next bounds are now an immediate consequence,*

$$\mathbb{E}\big(\sup_{[x]\in\mathcal{H}}|\overline{\nu_{[x]}}|^2-6H^2\big)_+\leqslant C[\frac{v}{n}\exp(\frac{-nH^2}{6v})+\frac{h^2}{n^2}\exp(\frac{-nH}{100h})] \tag{C.3}$$

$$\mathbb{P}\big(\sup_{[x]\in\mathcal{H}}|\overline{\nu_{[x]}}|^2\geqslant 6H^2\big)\leqslant 3\big[\exp\big(\frac{-nH^2}{400v}\big)+\exp\big(\frac{-nH}{200h}\big)\big] \tag{C.4}$$

In the sequel we will make use of the slightly simplified bounds (C.3) *and* (C.4) *rather than* (C.1) *and* (C.2). □

Remark C.0.2 *Introduce further the unit ball* $\mathbb{B}_m := \{[x]\in\Theta_{\overline{m}} : \|[x]\|_{l^2}\leqslant 1\}$ *contained in the linear subspace* $\Theta_{\overline{m}}=\overline{\mathrm{lin}}\big\{(\mathbb{1}_{\{s'=s\}})_{s'\in\mathbb{Z}}, |s|\in[\![1,m]\!]\big\}$.
Setting $\nu_{[x]}(Y)=\sum_{|s|\in[\![1,m]\!]}\overline{[x]}(s)\lambda^{-1}(s)e_s(-Y)$ *with* $\mathbb{E}\,\nu_{[x]}(Y)=\sum_{|s|\in[\![1,m]\!]}\overline{[x]}(s)\lambda^{-1}(s)\phi(s)$, *hence* $\overline{\nu_{[x]}}=\frac{1}{n}\sum_{p=1}^{n}\sum_{|s|\in[\![1,m]\!]}\overline{[x]}(s)\lambda^{-1}(s)(e_s(-Y_p)-\phi(s))$ *and we have*

$$\|\theta_{n,\overline{m}}-\theta^{\circ}_{\overline{m}}\|^2_{l^2}=\sup_{[x]\in\mathbb{B}_m}|\langle\theta_{n,\overline{m}}-\theta^{\circ}_{\overline{m}}|[x]\rangle_{l^2}|^2=\sup_{[x]\in\mathbb{B}_m}|\sum_{|s|\in[\![1,m]\!]}\lambda^{-1}(s)(\phi_n(s)-\phi(s))\overline{[x]}(s)|^2$$
$$=\sup_{[x]\in\mathbb{B}_m}|\sum_{|s|\in[\![1,m]\!]}\lambda^{-1}(s)\{\frac{1}{n}\sum_{i=1}^{n}(e_s(-Y_i)-\phi(s))\}\overline{[x]}(s)|^2=\sup_{[x]\in\mathbb{B}_m}|\overline{\nu_{[x]}}|^2.$$

Note that, the unit ball $\mathbb{B}_m$ *is not a countable set of functions, however, it contains a countable dense subset, say* $\mathcal{H}$, *since* l_2 *is separable, and it is straightforward to see that* $\sup_{[x]\in\mathbb{B}_m}|\overline{\nu_{[x]}}|^2=\sup_{[x]\in\mathcal{H}}|\overline{\nu_{[x]}}|^2$. □

C.1 Proofs for section 3.4.1

Proof of Lemma 3.4.1.
For $[x]\in\mathbb{B}_m$ setting

$$\nu_{[x]}(Y)=\sum_{|s|\in[\![1,m]\!]}\overline{[x]}(s)\lambda^{-1}(s)e_s(-Y)$$
$$\text{where }\mathbb{E}\,\nu_{[x]}(Y)=\sum_{|s|\in[\![1,m]\!]}\overline{[x]}(s)\lambda^{-1}(s)\phi(s)$$

we observe (see Remark C.0.2) that $\|\theta_{n,\overline{m}}-\theta^{\circ}_{\overline{m}}\|^2_{l^2}=\sup_{[x]\in\mathbb{B}_m}|\overline{\nu_{[x]}}|^2$. We intent to apply Lemma C.0.1. Therefore, we compute next quantities h, H, and v verifying the three inequalities required in Lemma C.0.1.
Consider h first:

$$\sup_{[x]\in\mathbb{B}_m}\sup_{y\in[0,1]}|\nu_{[x]}(y)|^2=\sup_{y\in[0,1]}\sum_{|s|\in[\![1,m]\!]}|\lambda(s)|^{-2}\,|e_s(y)|^2=2\sum_{s\in[\![1,m]\!]}\Lambda(s)$$
$$=2m\Lambda_{\circ}(m)\leqslant 2m\Lambda_+(m)=:h^2.$$

Next, find H. Notice that $\sup_{[x]\in\mathbb{B}_m}|\langle\theta_{n,\overline{m}}-\theta^{\circ}_{\overline{m}},[x]\rangle_{l^2}|^2=\sum_{|s|\in[\![1,m]\!]}\Lambda(|s|)\,|\phi_n(s)-\phi(s)|^2$.

As $\mathbb{E}\,|\phi_n(s)-\phi(s)|^2 = \frac{1}{n}(1-|\phi(s)|^2) \leqslant \frac{1}{n}$, we define

$$\mathbb{E}[\sup_{[x]\in\mathbb{B}_m} |\langle\theta_{n,\overline{m}} - \theta^\circ_{\overline{m}}\rangle^2_{l^2}] \leqslant 2m\Lambda_\circ(m)/n \leqslant \delta_\Lambda(m)2m\Lambda_+(m)/n = 2\Delta_\Lambda(m)/n =: H^2.$$

Finally, consider v. Given $[x]\in\mathbb{B}_m$ we observe with $\mathbb{E}[e_s(Y_1)e_{s'}(-Y_1)] = \phi(s'-s)$ that

$$\begin{aligned}\mathbb{E}\,|\nu_{[x]}(Y_1)|^2 &= \mathbb{E}\left|\sum\nolimits_{|s|\in[\![1,m]\!]} \overline{[x]}(s)\lambda^{-1}(s)e_s(-Y_1)\right|^2 \\ &= \sum_{|s|,|s'|\in[\![1,m]\!]} [x](s)\overline{[\lambda]}(s)^{-1}\,\mathbb{E}[e_s(Y_1)e_{s'}(-Y_1)]\lambda^{-1}(s')\,\overline{[x]}(s') \\ &= \sum_{|s|,|s'|\in[\![1,m]\!]} [x](s)\overline{[\lambda]}(s)^{-1}\phi(s'-s)\lambda^{-1}(s')\,\overline{[x]}(s') = \langle \mathcal{U}_k A\mathcal{U}_k\,[x],[x]\rangle_{l^2}\end{aligned}$$

defining the Hermitian and positive semi-definite matrix $A := (\overline{[\lambda]}(s)^{-1}\lambda^{-1}(s')\phi(s'-s))_{s,s'\in\mathbb{Z}}$ and the mapping $\mathcal{U}_k:\mathbb{C}^{\mathbb{Z}}\to\mathbb{C}^{\mathbb{Z}}$ with $z\mapsto\mathcal{U}_k z = (z_l\mathbb{1}_{\{|l|\in[\![1,m]\!]\}})_{l\in\mathbb{Z}}$. Obviously, $\mathcal{U}_k$ is an orthogonal projection from l^2 onto the linear subspace spanned by all l^2-sequences with support on the index-set $[\![-m,-1]\!]\cup[\![1,m]\!]$. Straightforward algebra shows $\sup_{[x]\in\mathbb{B}_m}\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_1)) \leqslant \sup_{[x]\in\mathbb{B}_m}\langle\mathcal{U}_k A\mathcal{U}_k\,[x],[x]\rangle_{l^2}$, hence

$$\sup_{[x]\in\mathbb{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_i)) \leqslant \sup_{[x]\in\mathbb{B}_m}\langle\mathcal{U}_kA\mathcal{U}_k[x],[x]\rangle_{l^2} = \sup_{[x]\in\mathbb{B}_m}\|\mathcal{U}_kA\mathcal{U}_k[x]\|_{l^2} \leqslant \|\mathcal{U}_kA\mathcal{U}_k\|_s.$$

where $\|M\|_s := \sup_{\|x\|_{l^2}\leqslant 1}\|Mx\|_{l^2}$ denotes the spectral-norm of a linear map $M: l^2\to l^2$. For a sequence $z\in(\mathbb{C}\backslash\{0\})^{\mathbb{Z}}$ let ∇_z and ∇_z^{-1} be the multiplication operator given by $\nabla_z\,x := (z(s)x(s))_{s\in\mathbb{Z}}$ and $\nabla_z^{-1} := \nabla_{z^{-1}}$, respectively. Clearly, we have $\mathcal{U}_kA\mathcal{U}_k = \mathcal{U}_k\,\nabla_\lambda^{-1}\,\mathcal{U}_k\mathcal{C}_\phi\mathcal{U}_k\,\nabla_{\overline{[\lambda]}}^{-1}\,\mathcal{U}_k$, where $\mathcal{C}_\phi := ([\phi]s-s')_{s,s'\in\mathbb{Z}}$. Consequently,

$$\begin{aligned}\sup_{[x]\in\mathbb{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_i)) &\leqslant \|\mathcal{U}_k\,\nabla_\lambda^{-1}\,\mathcal{U}_k\|_s\,\|\mathcal{C}_\phi\|_s\,\|\mathcal{U}_k\,\nabla_{\overline{[\lambda]}}^{-1}\,\mathcal{U}_k\|_s \\ &= \|\mathcal{U}_k\,\nabla_\lambda^{-1}\,\mathcal{U}_k\|_s^2\,\|\mathcal{C}_\phi\|_s.\end{aligned}$$

We have that $\|\mathcal{U}_k\,\nabla_\lambda^{-1}\,\mathcal{U}_k\|_s^2 = \max\{\Lambda(s), s\in[\![1,m]\!]\} = \Lambda_+(m)$. It remains to show the boundedness of $\|\mathcal{C}_\phi\|_s$. Keeping in mind that $(\mathcal{C}_\phi z)_k := \sum_{s\in\mathbb{Z}}\phi(s-k)z(s)$, $k\in\mathbb{Z}$, it is easily verified that $\|\mathcal{C}_\phi z\|^2_{l^2} \leqslant \|\phi\|^2_{l^1}\|z\|^2_{l^2}$ and hence $\|\mathcal{C}_\phi\|_s \leqslant \|\phi\|_{l^1}$, which implies

$$\sup_{[x]\in\mathbb{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_i)) \leqslant \|\phi\|_{l^1}\,\Lambda_+(m) \leqslant \|\phi\|_{l^1}\,\Lambda_+(m) =: v.$$

Replacing in Remark C.0.1 (C.3) and (C.4) the quantities h, H and v together with $\Delta_\Lambda(m) = \delta_\Lambda(m)m\Lambda_+(m)$ gives the assertion (i) and (ii). Setting $H := 2\mathfrak{R}^m_n(\theta^\circ,\Lambda) = 2[\mathfrak{b}^2_m(\theta^\circ)\vee\Delta_\Lambda(m)\,n^{-1}] \geqslant 2\Delta_\Lambda(m)\,n^{-1}$ and the quantities h as above v we obtain (iii), which completes the proof. □

C.2 Proofs for section 3.4.2

Proof of Lemma 3.4.2

In this part, let m and l be two positive integers with $l < m$. We have, for any t in $\mathbb{B}_{l,m}$

$$\langle t|\theta_n\rangle_{l^2} = n^{-1}\sum\nolimits_{p=1}^{n}\sum\nolimits_{l\leqslant|s|\leqslant m}(t(s)\overline{\lambda}(s)^{-1}\cdot e_s(-Y_p)) = n^{-1}\sum\nolimits_{p=1}^{n}\mathcal{F}^{-1}(t\overline{\lambda}^{-1})(-Y_p).$$

So we define for any t in $\mathbb{B}_{l,m}$ the functional $\nu_t := \sum_{l\leqslant|s|\leqslant m}(t(s)\overline{\lambda}(s)^{-1})e_s = \mathcal{F}^{-1}(t\overline{\lambda}^{-1})$ and we obtain $\overline{\nu}_t := \langle t|\theta_n - \theta^\circ\rangle_{l^2} = n^{-1}\sum_{p=1}^{n}(\nu_t(Y_p) - \mathbb{E}[\nu_t(Y_p)])$. Then, for any t in $\mathbb{B}_{l,m}$ and x in $\mathcal{L}^2$ we define $v_t(x) = w^{-1}\sum_{p=1}^{w}\nu_t(x_p)$, so we can write $n^{-1}\sum_{p=1}^{n}\nu_t(Y_p) = \frac{1}{2}\{r^{-1}\sum_{q=1}^{r}v_t(E_q) + r^{-1}\sum_{q=1}^{r}v_t(O_q)\}$, which gives

$$\begin{aligned}\langle t|\theta_n - \theta^\circ\rangle =& n^{-1}\sum\nolimits_{p=1}^{n}(\nu_t(Y_p) - \mathbb{E}[\nu_t(Y_p)])\\ =&\frac{1}{2}(\underbrace{r^{-1}\sum\nolimits_{q=1}^{r}(v_t(E_q) - \mathbb{E}[v_t(E_q)])}_{=:\overline{\nu}_t^e} + \underbrace{r^{-1}\sum\nolimits_{q=1}^{r}(v_t(O_q) - \mathbb{E}[v_t(O_q)])}_{=:\overline{\nu}_t^o})\end{aligned}$$

Similarly, we define for any t in $\mathbb{B}_{l,m}$ the quantities $\overline{\nu}_t^{e,\perp} := r^{-1}\sum_{q=1}^{r}(v_t(E_q^\perp) - \mathbb{E}[v_t(E_q^\perp)])$ and $\overline{\nu}_t^{o,\perp} := r^{-1}\sum_{q=1}^{r}(v_t(O_q^\perp) - \mathbb{E}[v_t(O_q^\perp)])$ which combined give $\overline{\nu}_t^\perp := \frac{1}{2}(\overline{\nu}_t^{e,\perp} + \overline{\nu}_t^{o,\perp})$. Consider first eq. (3.44).

$$\begin{aligned}\mathbb{E}\quad&[(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\langle t|\theta_n - \theta^\circ\rangle_{l^2}|^2 - C_n)_+] = \mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t|^2 - C_n)_+]\\ &\leqslant \mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{e,\perp}|^2 - C_n)_+] + \mathbb{E}[\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{e,\perp} - \overline{\nu}_t^e|^2] +\\ &\quad\mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{o,\perp}|^2 - C_n)_+] + \mathbb{E}[\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{o,\perp} - \overline{\nu}_t^o|^2]\\ &\leqslant 2\cdot\mathbb{E}[(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{e,\perp}|^2 - C_n)_+] + 2\cdot\mathbb{E}[\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{e,\perp} - \overline{\nu}_t^e|^2]\end{aligned}$$

which proves the statement.
Consider now eq. (3.45).

$$\begin{aligned}&\mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\langle t|\theta_n - \theta^\circ\rangle_{l^2}| \geqslant C_n) = \mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t| \geqslant C_n)\\ &\quad= \mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|(\overline{\nu}_t^e - \overline{\nu}_t^{e,\perp} + \overline{\nu}_t^{e,\perp} + \overline{\nu}_t^o - \overline{\nu}_t^{o,\perp} + \overline{\nu}_t^{o,\perp})/2| \geqslant C_n)\\ &\quad= \mathbb{P}(\{\sup\nolimits_{t\in\mathbb{B}_{l,m}}|(\overline{\nu}_t^e - \overline{\nu}_t^{e,\perp} + \overline{\nu}_t^{e,\perp} + \overline{\nu}_t^o - \overline{\nu}_t^{o,\perp} + \overline{\nu}_t^{o,\perp})/2| \geqslant C_n\}\\ &\quad\cap_{q\in[\![1,r]\!]}(\{E_q^\perp = E_q\}\cap\{E_q^\perp = E_q\}))\\ &\quad+ \mathbb{P}(\{\sup\nolimits_{t\in\mathbb{B}_{l,m}}|(\overline{\nu}_t^e - \overline{\nu}_t^{e,\perp} + \overline{\nu}_t^{e,\perp} + \overline{\nu}_t^o - \overline{\nu}_t^{o,\perp} + \overline{\nu}_t^{o,\perp})/2| \geqslant C_n\}\\ &\quad\cap(\{\exists q\in[\![1,r]\!], E_q^\perp \neq E_q\}\cup\{\exists q\in[\![1,r]\!], O_q^\perp \neq O_q\}))\\ &\quad\leqslant \mathbb{P}(\sup\nolimits_{t\in\mathbb{B}_{l,m}}|\overline{\nu}_t^{e,\perp}| \geqslant C_n) + 2\sum_{q=1}^{r}\mathbb{P}(\{E_q^\perp \neq E_q\})\end{aligned}$$

Which completes the proof. □

PROOF OF LEMMA 3.4.3

Let be m and l in $\mathbb{N}$ with $m \leqslant l$ throughout this proof. Recall that we want to bound $\mathbb{E}[(\sup_{t\in\mathbb{B}_{m,l}} |\overline{\nu}_t^{e,\perp}|^2 - 6H^2)_+]$ and $\mathbb{P}(\sup_{t\in\mathbb{B}_{m,l}} |\overline{\nu}_t^{e,\perp}| \geqslant 6H^2)$, where, for any t in $\mathbb{B}_{\underline{m},\bar{l}}$

$$\overline{\nu}_t^{e,\perp} = r^{-1}\sum_{q=1}^{r}(v_t(E_q^\perp) - \mathbb{E}[v_t(E_q^\perp)]); \quad v_t(E_q^\perp) = w^{-1}\sum_{p=1}^{w}\nu_t(E_{q,p}^\perp);$$
$$\nu_t(E_{q,p}^\perp) = \sum_{m\leqslant|s|\leqslant l}(t(s)\overline{\lambda}(s)^{-1}e_s(E_{q,p}^\perp)).$$

and H is such that $H^2 \geqslant n^{-1}\Lambda_+(l)(l-m+1)(\psi_m+1)$.

We will use Talagrand's inequality (Lemma C.0.1). Recall that to do so, we have to exhibit three real numbers h, H and v verifying:

$$\sup_{t\in\mathbb{B}_{\underline{m},\bar{l}}}\sup_{y\in[0,1]^w}|v_t(y)| \leqslant h; \quad \mathbb{E}[\sup_{t\in\mathbb{B}_{\underline{m},\bar{l}}}|\overline{\nu}_t^{e,\perp}|] \leqslant H;$$
$$\sup_{t\in\mathbb{B}_{\underline{m},\bar{l}}} w^{-1}\sum_{p=1}^{w}\mathbb{V}_{\theta^\circ}[\nu_t(E_{q,p}^\perp)] \leqslant v.$$

We start with h which gives us

$$\sup_{t\in\mathbb{B}_{\underline{m},\bar{l}}}\sup_{y\in[0,1]^w}|v_t(y)|^2 = \sup_{t\in\mathbb{B}_{\underline{m},\bar{l}}}\sup_{y\in[0,1]^w}|w^{-1}\sum_{p=1}^{w}\nu_t(y_p)|^2$$
$$= \sup_{y\in[0,1]^w}\sum_{s=m}^{l}|w^{-1}\sum_{p=1}^{w}e_s(y_p)|^2\Lambda(s) \leqslant \sum_{m\leqslant|s|\leqslant l}\Lambda(s).$$

Hence we define $h^2 := \delta_{m,l}^\star \geqslant \sum_{m\leqslant|s|\leqslant l}\Lambda(s)$.

Considering H^2, we define the following objects: $\overline{e}_s^\perp := (r\cdot w)^{-1}\sum_{q=1}^{r}\sum_{p=1}^{w}e_s(E_{q,p}^\perp)$ and $\overline{e}^\perp = (\overline{e}_s(E_q^\perp))_{s\in\mathbb{Z}}$. We first replace $\overline{\nu}_t^{e,\perp}$, then $v_t(E_q^\perp)$ and finally $\nu_t(E_{q,p}^\perp)$ by their respective definition; using Fubini theorem, a scalar product appears and we use $\mathbb{E}[e_s(E_{q,1}^\perp)] = \phi(s)$ as $E_{q,1}^\perp \sim \mathbb{P}_\phi$; Riesz representation theorem allows to get rid of the supremum, we then use the linearity of expectation and independence of the blocs and finally we use Lemma 1.2.3 in the last line with $K = \lfloor\delta_\Lambda(m)\sqrt{(l-m+1)}(4\gamma_g)^{-1}\rfloor$ which tends to infinity and is smaller than $w-1$ and conclude (for w large enough):

$$\mathbb{E}[\sup_{t\in\mathbb{B}_{\underline{m},l}}|\overline{\nu}_t^{e,\perp}|^2] \leqslant 8n^{-1}\Lambda_+(l)(l-m+1)\delta_\Lambda(m).$$

So we set $H^2 \geqslant 8n^{-1}\Lambda_+(l)(l-m+1)\delta_\Lambda(m)$.

Finally we control v. Using Lemma 1.2.1 and Cauchy-Schwarz inequality, we have

$$\sup_{[x]\in\mathbb{B}_{l,m}} r^{-1}\sum_{q=1}^{r}\mathbb{V}[v_{[x]}(E_q^\perp)] = \sup_{[x]\in\mathbb{B}_{l,m}} w^{-2}\,\mathbb{V}[\sum_{p=1}^{w}\nu_{[x]}(E_{1,p}^\perp)]$$
$$\leqslant 4w^{-1}\sup_{[x]\in\mathbb{B}_{l,m}}\mathbb{E}[|\nu_{[x]}(E_{q,0}^\perp)|^2 b(E_{q,0}^\perp)] \leqslant 4w^{-1}\sup_{[x]\in\mathbb{B}_{l,m}}\sqrt{\mathbb{E}[|\nu_{[x]}(E_{q,0}^\perp)|^2]\|\nu_{[x]}\|_\infty\,\mathbb{E}[b(E_{q,0}^\perp)]} \tag{C.5}$$

as we have already proven in appendix C.1 we have

$$\mathbb{E}[|\nu_{[x]}(E^{\perp}_{q,0})|^2] \leqslant \|\phi\|_{l^1}\Lambda_+(m), \quad \text{and} \quad \sup_{[x]\in\mathbb{B}_{l,m}} \|\nu_{[x]}\|_\infty \leqslant \sum_{s=l}^{m} \Lambda(s). \tag{C.6}$$

Combining C.5, and C.6, we obtain

$$\sup_{[x]\in\mathbb{B}_{l,m}} r^{-1}\sum_{q=1}^{r} \mathbb{V}[v_{[x]}(E^{\perp}_q)] \leqslant 4w^{-1}\sqrt{m}\Lambda_+(m)\sqrt{2\|\phi\|_{l^1}\sum_{p=1}^{\infty}(p+1)\beta_p} =: v.$$

Using Talagrand's inequality gives us the result. □

Proof of Lemma 3.4.5

Both inequalities are verified using $\mathbb{P}(E_q \neq E^{\perp}_q) \leqslant \beta_w$ and, as it was proven in appendix C.1, $\|v_t\|^2_\infty \leqslant \sum_{m\leqslant|s|\leqslant l}\Lambda(s)$.

Consider first eq. (3.51), that is to say

$$\mathbb{E}[\sup_{t\in\mathbb{B}_{m,l}}|\overline{\nu}^{e,\perp}_t - \overline{\nu}^{e}_t|^2] \leqslant 2r\beta_s\sum_{m\leqslant|s|\leqslant l}\Lambda(s).$$

To prove it we define the sequence of events for any two integers m and l with $m \leqslant l$; and t in $\mathbb{B}_{m,l}$ let be $\Omega_{r,s} := \bigcap_{q\in[\![1,r]\!]}\{E^{\perp}_q = E_q\}$. Notice that $\mathbb{P}(\Omega_{r,s}) \geqslant \sum_{q\in[\![1,r]\!]}\mathbb{P}_{\theta^\circ}(E^{\perp}_q = E_q) - r + 1 \geqslant r(1-\beta_s) - r + 1 \geqslant \max\{1 - r\beta_s, 0\}$. Then, we have

$$\mathbb{E}[\sup_{t\in\mathbb{B}_{m,l}}|\overline{\nu}^{e,\perp}_t - \overline{\nu}^{e}_t|^2] \leqslant 2\|\nu_t\|^2_\infty\,\mathbb{P}[\Omega^c_{r,s}] \leqslant 2\|\nu_t\|^2_\infty r\beta_s \leqslant 2r\beta_s\sum_{m\leqslant|s|\leqslant l}\Lambda(s);$$

which proofs the first statement. □

C.3 Proofs for section 3.4.3

Proof of Lemma 3.4.6.
For $[x] \in \mathbb{B}_m$ setting

$$\nu_{[x]}(Y) = \sum_{|s|\in[\![1,m]\!]} \overline{[x]}(s)\lambda^{+}_{n_\lambda}(s)e_s(-Y)$$

$$\text{where } \mathbb{E}_{Y|\varepsilon}\,\nu_{[x]}(Y) = \sum_{|s|\in[\![1,m]\!]} \overline{[x]}(s)\lambda^{+}_{n_\lambda}(s)\phi(s)$$

we observe (see Remark C.0.1) that $\|\theta_{n,n_\lambda,\overline{m}} - \breve{\theta}_{\overline{m}}\|^2_{l^2} = \sup_{[x]\in\mathbb{B}_m}|\overline{\nu_{[x]}}|^2$. We intent to apply Lemma C.0.1. Therefore, we compute next quantities h, H, and v verifying the three inequalities required in Lemma C.0.1.

Consider h first:

$$\sup_{[x]\in\mathbb{B}_m}\sup_{y\in[0,1]}|\nu_{[x]}(y)|^2=\sup_{y\in[0,1]}\sum_{|s|\in\llbracket 1,m\rrbracket}|\lambda^+_{n_\lambda}(s)|^2\,|e_s(y)|^2$$
$$=2\sum_{s\in\llbracket 1,m\rrbracket}\widehat{\Lambda}(s)=2m\Lambda_\circ(m)\leqslant 2m\widehat{\Lambda}_+(m)=:h^2.$$

Next, find H. Notice that $\sup_{[x]\in\mathbb{B}_m}|\langle\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}},[x]\rangle_{l^2}|^2=\sum_{|s|\in\llbracket 1,m\rrbracket}\widehat{\Lambda}|s|\,|\phi_n(s)-\phi(s)|^2$. As $\mathbb{E}_{Y|\varepsilon}|\phi_n(s)-\phi(s)|^2=\frac{1}{n}(1-|\phi(s)|^2)\leqslant\frac{1}{n}$, we define

$$\mathbb{E}_{Y|\varepsilon}[\sup_{[x]\in\mathbb{B}_m}|\langle\theta_{n,n_\lambda,\overline{m}}-\breve{\theta}_{\overline{m}}\rangle^2_{l^2}]\leqslant 2m\Lambda_\circ(m)/n\leqslant\delta_{\widehat{\Lambda}}(m)2m\widehat{\Lambda}_+(m)/n=2\Delta_{\widehat{\Lambda}}(m)/n=:H^2.$$

Finally, consider v. Given $[x]\in\mathbb{B}_m$ we observe with

$$\mathbb{E}_{Y|\varepsilon}[e_s(Y_1)e_{s'}(-Y_1)]=\mathbb{E}[e_s(Y_1)e_{s'}(-Y_1)]=\phi(s'-s)$$

that

$$\mathbb{E}_{Y|\varepsilon}|\nu_{[x]}(Y_1)|^2=\mathbb{E}_{Y|\varepsilon}\left|\sum\nolimits_{|s|\in\llbracket 1,m\rrbracket}\overline{[x]}(s)\lambda^+_{n_\lambda}(s)e_s(-Y_1)\right|^2$$
$$=\sum\nolimits_{|s|,|s'|\in\llbracket 1,m\rrbracket}[x](s)\overline{\lambda^+_{n_\lambda}}(s)\,\mathbb{E}[e_s(Y_1)e_{s'}(-Y_1)]\lambda^+_{n_\lambda}(s')\,\overline{[x]}(s')$$
$$=\sum_{|s|,|s'|\in\llbracket 1,m\rrbracket}[x](s)\overline{\lambda^+_{n_\lambda}}(s)\phi(s'-s)\lambda^+_{n_\lambda}(s')\,\overline{[x]}(s')=\langle\mathcal{U}_k\widehat{A}\mathcal{U}_k\,[x],[x]\rangle_{l^2}$$

defining the Hermitian and positive semi-definite matrix $\widehat{A}:=(\overline{\lambda^+_{n_\lambda}}(s)\lambda^+_{n_\lambda}(s')\phi(s'-s))_{s,s'\in\mathbb{Z}}$ and the mapping $\mathcal{U}_k:\mathbb{C}^{\mathbb{Z}}\to\mathbb{C}^{\mathbb{Z}}$ with $z\mapsto\mathcal{U}_kz=(z_l\mathbb{1}_{\{|l|\in\llbracket 1,m\rrbracket\}})_{l\in\mathbb{Z}}$. Obviously, $\mathcal{U}_k$ is an orthogonal projection from l^2 onto the linear subspace spanned by all l^2-sequences with support on the index-set $\llbracket -m,-1\rrbracket\cup\llbracket 1,m\rrbracket$. Straightforward algebra shows $\sup_{[x]\in\mathbb{B}_m}\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_1))\leqslant\sup_{[x]\in\mathbb{B}_m}\langle\mathcal{U}_k\widehat{A}\mathcal{U}_k\,[x],[x]\rangle_{l^2}$, hence

$$\sup_{[x]\in\mathbb{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_i))\leqslant\sup_{[x]\in\mathbb{B}_m}\langle\mathcal{U}_k\widehat{A}\mathcal{U}_k[x],[x]\rangle_{l^2}=\sup_{[x]\in\mathbb{B}_m}\|\mathcal{U}_k\widehat{A}\mathcal{U}_k[x]\|_{l^2}\leqslant\|\mathcal{U}_k\widehat{A}\mathcal{U}_k\|_s.$$

where $\|M\|_s:=\sup_{\|x\|_{l^2}\leqslant 1}\|Mx\|_{l^2}$ denotes the spectral-norm of a linear map $M:l^2\to l^2$. For a sequence $z\in(\mathbb{C}\backslash\{0\})^{\mathbb{Z}}$ let ∇_z be the multiplication operator given by $\nabla_z x:=(z(s)x(s))_{s\in\mathbb{Z}}$. Clearly, we have $\mathcal{U}_k\widehat{A}\mathcal{U}_k=\mathcal{U}_k\nabla_{\lambda^+_{n_\lambda}(s)}\mathcal{U}_k\mathcal{C}_\phi\mathcal{U}_k\nabla_{\overline{\lambda^+_{n_\lambda}}[(s)]}\mathcal{U}_k$, where $\mathcal{C}_\phi:=([\phi]s-s')_{s,s'\in\mathbb{Z}}$. Consequently,

$$\sup_{[x]\in\mathbb{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_i))\leqslant\|\mathcal{U}_k\nabla_{\lambda^+_{n_\lambda}(s)}\mathcal{U}_k\|_s\,\|\mathcal{C}_\phi\|_s\,\|\mathcal{U}_k\nabla_{\overline{\lambda^+_{n_\lambda}}[(s)]}\mathcal{U}_k\|_s$$
$$=\|\mathcal{U}_k\nabla_{\lambda^+_{n_\lambda}(s)}\mathcal{U}_k\|_s^2\,\|\mathcal{C}_\phi\|_s.$$

We have that $\|\mathcal{U}_k\nabla_{\lambda^+_{n_\lambda}(s)}\mathcal{U}_k\|_s^2=\max\{\widehat{\Lambda}(s),s\in\llbracket 1,m\rrbracket\}=\widehat{\Lambda}_+(m)$. It remains to show the boundedness of $\|\mathcal{C}_\phi\|_s$. Keeping in mind that $(\mathcal{C}_\phi z)_k := \sum_{s\in\mathbb{Z}} \phi(s-k)z(s)$, $k \in \mathbb{Z}$, it is easily verified that $\|\mathcal{C}_\phi z\|_{l^2}^2 \leqslant \|\phi\|_{l^1}^2 \|z\|_{l^2}^2$ and hence $\|\mathcal{C}_\phi\|_s \leqslant \|\phi\|_{l^1}$, which implies

$$\sup_{[x]\in\mathbb{B}_m} \tfrac{1}{n} \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\nu_{[x]}(Y_i)) \leqslant \|\phi\|_{l^1}\, \widehat{\Lambda}_+(m) =: v.$$

Replacing in Remark C.0.1 (C.3) and (C.4) the quantities h, H and v together with $\Delta_{\widehat{\Lambda}}(m) = \delta_{\widehat{\Lambda}}(m) m \widehat{\Lambda}_+(m)$ gives the assertion (i) and (ii). Setting $H := 2\mathfrak{R}_n^m \theta^\circ, \widehat{\Lambda} = 2[\mathfrak{b}_m^2(\theta^\circ) \vee \Delta_{\widehat{\Lambda}}(m)\, n^{-1}] \geqslant 2\Delta_{\widehat{\Lambda}}(m)\, n^{-1}$ and the quantities h as above v we obtain (iii), which completes the proof. □

Proof of Lemma 3.4.7.

The assertion follows directly from Hoeffding's inequality. Indeed, setting $\widetilde{X}_i := e_s(-\varepsilon_i) - \lambda(s)$, $i \in [\![1, n_\lambda]\!]$, obviously $\widetilde{X}_1, \dots, \widetilde{X}_{n_\lambda}$ are iid. with mean zero and $|\widetilde{X}_i| \leqslant d = 2$, hence

$$\begin{aligned}
\mathbb{P}\left(|\lambda_{n_\lambda}(s)/\lambda(s) - 1| > 1/3\right) &= \mathbb{P}\left(|\lambda_{n_\lambda}(s) - \lambda(s)| > |\lambda(s)|/3\right) \\
&= \mathbb{P}\left(|\sum_{i=1}^{n} \widetilde{X}_i| > n_\lambda |\lambda(s)|/3\right) \leqslant 2\exp(-\tfrac{(n_\lambda|\lambda(s)|/3)^2}{2d^2 n_\lambda}) = 2\exp(-\tfrac{n_\lambda |\lambda(s)|^2}{72}).
\end{aligned}$$

□

Bibliography

Asin, N. and Johannes, J. (2016). Adaptive non-parametric instrumental regression in the presence of dependence. *Annals of Economics and Statistics*.

Asin, N. and Johannes, J. (2017). Adaptive nonparametric estimation in the presence of dependence. *Journal of Nonparametric Statistics*, 29:1–37.

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability theory and related fields*, 113(3):301–413.

Bigot, J. and van Bellgem, S. (2009). Log-density deconvolution by wavelet thresholding. *Scandinavian Journal of Statistics*, 36:749–763.

Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons, second edition.

Birgé, L. (2001). An alternative point of view on lepski's method. In Monogr., editor, *State of the art in probability and statistics*, volume 36 of *IMS Lecture Notes*, pages 113–133. (Leiden 1999).

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *The Annals of Statistics*, 23(1):11–29.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2:107–144.

Bunke, O. and Johannes, J. (2005). Selfinformative limits of bayes estimates and generalized maximum likelihood. *Statistics*, 39(6):483–502.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83:1184–1186.

Cavalier, L. and Hengartner, N. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361.

Cavalier, L. and Raimondo, M. (2007). Wavelet deconvolution with noisy eigenvalues. *IEEE Transactions on signal processing*, 55(6):2414–2424.

Cochran, W. W., Mouritsen, H., and Wikelski, M. (2004). Migrating songbirds recalibrate their magnetic compass daily from twylight cues. *Science*, 304:405.

Comte, F. and Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM: Probability and Statistics*, 6:211–238.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3):431–452.

Comte, F. and Taupin, M.-L. (2003). Adaptive density deconvolution for circular data. Prépublication map5 2003-10, Université Paris Descartes.

Curray, J. R. (1956). The analysis of two-dimensional orientation data. *The Journal of Geology*, 64(2):117–131.

Devroye, L. (1989). Consistent deconvolution in density estimation. *Canadian Journal of Statistics*, 17:235–239.

Doob, J. L. (1949). Application of the theory of martingales. *Le calcul des probabilites et ses applications*, pages 23–27.

Dunker, F., Florens, J.-P., Hohage, T., Johannes, J., and Mammen, E. (2014). Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression. *Journal of Econometrics*, 178:444–455.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92:526–535.

Engl, H. W., Kunisch, K., and Neubauer, A. (1989). Convergence rates for tikhonov regularisation of non-linear ill-posed problems. *Inverse problems*, 5(4):523.

Engle, H., Hanke, M., and Neubauer, A. (1996). Regularization of inverse problems, volume 375 of mathematics and its applications.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, 19:1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canadian Journal of Statistics*, 20:734–747.

Fan, J. and Koo, J. Y. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory*, 48:734–747.

Fisher, N. (1993). *Statistical analysis of circular data*. Cambridge University Press.

Ghosal, S., Ghosh, J. K., and van Der Vaart, A. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press.

Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in lr-metrics, 1⩽ r⩽∞. *The Annals of Statistics*, 39(6):2883–2911.

Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli*, 5:907–925.

Goldenshluger, A. (2000). Density deconvolution in the circular structural model. *Journal of Multivariate Analysis*, 81:360–375.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39:1608–1632.

Hadamard, J. (1902). Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52.

Heinz, E. (1951). Beiträge zur Störungstheorie der Spektralzerlegung. *Mathematische Annalen*, 123:415–438.

Johannes, J. and Schenk, R. (2013). On rate optimal local estimation in functional linear regression. *Electronic Journal of Statistics*, 7:191–216.

Johannes, J., Schenk, R., and Simoni, A. (2014). Adaptive bayesian estimation in gaussian sequence space models. *Contributions in infinite-dimensional statistics and related topics*, pages 167–172.

Johannes, J. and Schwarz, M. (2013a). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli*, 19(5A):1576–1611.

Johannes, J. and Schwarz, M. (2013b). Adaptive gaussian inverse regression with partially unknown operator. *Communications in Statistics - Theory and Methods*, 42(7):1343–1362.

Kim, P. T. and Koo, J. Y. (2002). Optimal spherical deconvolution. *Journal of Multivariate Analysis*, 80:21–42.

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077.

Knapik, B. T., van der Vaart, A. W., and van Zanten, J. H. (2011). Bayesian inverse problems with gaussian priors. *The Annals of Statistics*, 39(5):2626–2657.

Koo, J. Y. and Park, B. U. (1996). B-spline deconvolution based on the EM algorithm. *Journal of Statistical Computation and Simulation.*, 54:275–288.

Laurent, B., Loubes, J., and Marteau, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6:91–122.

Liptser, R. S. and Shiryaev, A. N. (2013). *Statistics of random processes: I. General theory*, volume 5. Springer Science & Business Media.

Mardia, K. (1972). *Statistics of directional data.* Probability and Mathematical Statistics. A Series of Monographs and Textbooks. Vol. 13. Academic Press.

Massart, P. (2007). Concentration inequalities and model selection.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics.* Lecture Notes in Statistics 193. Berlin: Springer.

Mendelsohn, J. and Rice, J. (1982). Deconvolution of micro-fluorometric histograms with B-splines. *Journal of the American Statistical Association*, 77:748–753.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7:307–330.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, 27:2033–2053.

Petrov, V. V. (1995). *Limit theorems of probability theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press., Oxford, 4. edition.

Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer.

Rigollet, P. and Tsybakov, A. B. (2007). Linear and convex aggregation of density estimators. *Mathematical Methods of Statistics*, 16(3):260–280.

Schwartz, L. (1965). On bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4(1):10–26.

Stefanski, L. A. (1990). Rate of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters*, 9:229–235.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563.

Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probability theory and related fields*, 107(4):467–492.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Annals of Statistics*, 18:806–831.