Adaptive Bayesian estimation and self informative limit in indirect Gaussian sequence space models

Jan Johannes   Xavier Loizeau

1 Introduction

1.1 The inverse Gaussian sequence space model and notations in use

In the following article, we focus on the inverse gaussian sequence space model.
This model is defined as follow: let $\left(\theta_j^\circ\right)_{j\in\mathbb{N}^\star}$ be a square summable sequence of unknown real numbers which we want to estimate ; $(\lambda_j)_{j\in\mathbb{N}^\star}$, a sequence of known real numbers to be interpreted as Eigen values of a known diagonal linear operator, n a natural integer which represents the noise level, and $(\boldsymbol{\xi}_j)_{j\in\mathbb{N}^\star}$ be a sequence of independent and identically distributed random variables (iid) with normal standard distribution (for all j, $\xi_j \sim_{iid} \mathcal{N}(0,1)$). Then, we observe the sequence $\left(Y_j^n\right)_{j\in\mathbb{N}^\star}$, defined by :

$$\forall j \in \mathbb{N}^\star, \quad Y_j^n = \lambda_j \cdot \theta_j^\circ + \frac{1}{\sqrt{n}}\boldsymbol{\xi}_j.$$

Note that the elements of the sequence $\left(Y_j^n\right)_{j\in\mathbb{N}^\star}$ sequence are not identically distributed and that for all j in $\mathbb{N}^\star$, Y_j^n follows $\mathcal{N}(\lambda_j \cdot \theta_j^\circ, \frac{1}{n})$, however, they are independent.
In the following, we will denote $\mathbb{P}_{\theta^\circ}^n$ the distribution of Y^n and $\mathbb{E}_{\theta^\circ}^n$ the expectation under this distribution. Obviously, these two objects depend on n and θ°.

1.2 Metrics and Sobolev's ellipsoids

We will from now on note Θ the set of square summable sequences of real numbers: $\Theta := \left\{\theta \in \mathbb{R}^{\mathbb{N}^\star} : \sum_{j\in\mathbb{N}^\star} \theta_j^2 < \infty\right\}$.
A natural norm on Θ is the L^2-norm, denoted $\|\cdot\|$ and defined as the application which associates to any θ in Θ the positive, finite, real number $\|\theta\| = \sqrt{\sum_{j\in\mathbb{N}^\star} \theta_j^2}$.
Alternatively one can consider weighted norms. For any positive, non-increasing sequence of real numbers a, define the set $\Theta_a := \left\{\theta \in \mathbb{R}^{\mathbb{N}^\star} : \sum_{j\in\mathbb{N}^\star} \theta_j^2/a_j < \infty\right\}$. We define the a-weighted norm over Θ_a as the application which to any θ in Θ_a associates $\|\theta\|_a = \sqrt{\sum_{j\in\mathbb{N}^\star} \theta_j^2/a_j}$.

Common sets over which to consider minimax optimality (see SECTION 2.2 and SECTION 3.2) are Sobolev's ellipsoids. Define a positive non-increasing sequence $(\mathfrak{a}_j)_{j\in\mathbb{N}^\star} \in$

$\mathbb{R}_+^{\mathbb{N}^\star}$ such that $\mathfrak{a}_1 = 1$ and $\lim_{j\to\infty} \mathfrak{a}_j = 0$ and r a real constant. Then we have $\Theta_{\mathfrak{a}}(r) = \{\theta \in \Theta_{\mathfrak{a}} : \|\theta\|_{\mathfrak{a}} \leq r\}$.

1.3 Definitions and set of assumptions

We introduce further objects which appear to have a meaningful role subsequently.

DEFINITION 1.1 We define for all m in $\mathbb{N}^\star$

$$\begin{aligned}
\mathfrak{b}_m &:= \sum_{j>m} \left(\theta_j^\circ\right)^2; & \Lambda_m &:= \frac{1}{\lambda_m^2}; & \overline{\Lambda}_m &:= \frac{1}{m}\sum_{j=1}^{m} \Lambda_j;\\
\Phi_n^m &:= \left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]; & m_n^\circ &:= m_n^\circ(\theta^\circ,\lambda) = \arg\min_{m\in\mathbb{N}^\star}\left\{\Phi_n^m\right\}; & \Phi_n^\circ &:= \Phi_n^\circ(\theta^\circ,\lambda) = \Phi_n^{m_n^\circ};\\
\Psi_n^m &:= \left[\mathfrak{a}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]; & m_n^\star &:= m_n^\star(\mathfrak{a},\lambda) = \arg\min_{m\in\mathbb{N}^\star}\left\{\Psi_n^m\right\}; & \Psi_n^\star &:= \Psi_n^\star(\mathfrak{a},\lambda) = \Psi_n^{m_n^\star}.
\end{aligned}$$

We will see in SECTION 2.1 that $\mathfrak{b}_m$ is the bias for a projection estimator and $m \cdot \overline{\Lambda}_m/n$ is its variance, moreover, we show in SECTION 3.1 and SECTION 3.2 that Φ_n° and $\Psi_n^\star$ are respectively oracle and minimax optimal contraction rates.
To prove results about the adaptive methods we study in SECTION 3.3 and SECTION 4.3, the following assumptions on the model are required.

ASSUMPTION 1.1 Suppose that λ is monotonically and polynomially decreasing, that is, there exist c in $[1,\infty[$ and a in $\mathbb{R}_+$ such that

$$\forall j \in \mathbb{N}^\star, \quad \frac{1}{c} j^{-a} \leq \lambda_j \leq c j^{-a}.$$

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

$$\sup_{n\in\mathbb{N}^\star} \frac{m_n\Lambda_{m_n}}{n\Phi_n^{m_n}} \leq \sup_{n\in\mathbb{N}^\star} \Lambda_{m_n}/\overline{\Lambda}_{m_n} \leq L.$$

ASSUMPTION 1.2 Let θ° and λ be such that there exists n° in $\mathbb{N}^\star$

$$0 < \kappa^\circ := \kappa^\circ(\theta^\circ,\lambda) := \inf_{n\geq n^\circ}\left\{(\Phi_n^\circ)^{-1}\left[\mathfrak{b}_{m_n^\circ} \wedge \frac{m_n^\circ\overline{\Lambda}_{m_n^\circ}}{n}\right]\right\} \leq 1$$

ASSUMPTION 1.3 Let $\mathfrak{a}$ and λ be sequences such that there exists $n^\star$ in $\mathbb{N}^\star$

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

Unfortunately, these assumptions, though only required in the adaptive case, rule out the possibility of a severely ill-posed problem as well as the possibility of a "parametric model" (in the sense of all entries of θ° canceling after a certain index).

1.4 Main results of this article

An important feature of this article is showing that Φ_n° and $\Psi_n^\star$ are respectively oracle (for a family of sieve priors) and minimax (over Sobolev's ellipsoids) optimal contraction rates as defined in SECTION 3.1 and SECTION 3.2. In SECTION 4.3, we exhibit a family of fully data-driven Bayesian methods which, under some hypotheses reach these rates of contraction and which posterior means are optimal estimators as formulated in SECTION 2.1 and SECTION 2.2. In addition, the family of Bayesian methods we consider is indexed by a so-called iteration parameter and we show that, if one lets this iteration parameter tend to infinity, then, the posterior distribution is concentrating on the well known projection estimator with model selection by penalised contrast. By doing so, we are able to proof optimality of this estimator in a novel way.

The main novelties of this article reside on the general formulation we adopt for Bayesian optimality; the influence on the contraction rate of iteration parameter we use to generate the family of Bayesian methods (first introduced in BUNKE AND JOHANNES (2005)) hasn't been studied yet, up to our knowledge and the strategy to proof the rates of contraction in the non-adaptive case, though simple, seems new.
The study of contraction rates under the L^2-norm remains marginal despite major contributions such as JJASRS, Nickl,... As well as the absence of a log-loss term. However, for those two points this article could not push further the results obtained in JOHANNES ET AL. (2016), therefore, in the adaptive case, we formulated some of their hypotheses on a stronger way to simplify our proofs, however, a similar set of assumptions could be used.

2 Frequentist inference

We briefly introduce here the principles of oracle and minimax optimal rates of convergence by taking the example of a family of projections estimators which appear to have an important meaning in SECTION 4.
From a frequentist point of view, one would observe a realisation Y^n from $\mathbb{P}^n_{\theta^\circ}$ for some θ° in Θ and they then would like to infer on θ°. To do so, the estimation procedure consists in considering an estimator, that is to say a mapping from the space of observation to the parameter space, generally depending on n, and proving that this application fulfils optimality conditions such as minimising a risk.
In the SECTION 4.3 we give results about the well known rate of convergence in L^2 norm. However it makes sense to remind here the definition of the rate of convergence in distribution which plays an important role in SECTION 3.1 and SECTION 3.2 to define rates of contraction.

2.1 Oracle optimality

For any j in $\mathbb{N}^\star$, a natural estimator for θ_j° is $\overline{\theta}_j := Y_j/\lambda_j$ as it is unbiased ($\mathbb{E}^n_{\theta^\circ}\left[\overline{\theta}_j\right] = \theta_j^\circ$) and its variance is $\mathbb{V}^n_{\theta^\circ}\left[\overline{\theta}_j\right] = \Lambda_j/n$.
In this context, by projection estimators, one refers to the family of estimators of θ° defined by $\mathcal{F} := \left\{\overline{\theta}^m := \left(\overline{\theta}^m_j\right)_{j\in\mathbb{N}^\star} = \left(\overline{\theta}_j \mathbb{1}_{j\leq m}\right)_{j\in\mathbb{N}^\star} m \in \mathbb{N}^\star\right\}$.

We say that, for any θ° in Θ, the sequence Φ_n°, as defined in DEFINITION 1.1, is the oracle optimal rate of convergence in probability for the family of estimators $\mathcal{F}$ at θ° because, for any increasing (arbitrarily slowly) and unbounded sequence $(c_n)_{n\in\mathbb{N}^\star}$

$$\lim_{n\to\infty} \sup_{\overline{\theta}^{m_n}\in\mathcal{F}} \mathbb{P}_{\theta^\circ}^n\left[d^2\left(\theta^\circ,\overline{\theta}^{m_n}\right)\leq c_n^{-1}\Phi_n^\circ\right]=0;$$

and, with m_n° as in DEFINITION 1.1,

$$\lim_{n\to\infty} \mathbb{P}_{\theta^\circ}^n\left[d^2\left(\theta^\circ,\overline{\theta}^{m_n^\circ}\right)\leq c_n\cdot\Phi_n^\circ\right]=1.$$

We therefore call $\overline{\theta}^{m_n^\circ}$ oracle optimal. One should notice, though, that m_n° depends on θ° and is hence, not at hand in practice and one should define a data-driven way to select this parameter which conserves this optimality property.

2.2 Minimax optimality

Another form of optimality gathering a lot of interest is the minimax optimality.
In a frequentist framework, $\Psi_n^\star$, as defined in DEFINITION 1.1, is called minimax optimal convergence rate over $\Theta_{\mathfrak{a}}(r)$ as, for any increasing unbounded sequence $(c_n)_{n\in\mathbb{N}}$ and with $m_n^\star$ as in DEFINITION 1.1,

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{\widehat{\theta}} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}(r)} \mathbb{P}_{\theta^\circ}^n\left[d^2(\theta^\circ,\widehat{\theta})\leq c_n^{-1}\Psi_n^\star\right] &= 0;\\
&\lim_{n\to\infty} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}(r)} \mathbb{P}_{\theta^\circ}^n\left[d^2(\theta^\circ,\overline{\theta}^{m_n^\star})\leq c_n\Psi_n^\star\right] &= 1;
\end{aligned}$$

where $\sup_{\widehat{\theta}}$ is taken over all possible estimators. Any estimator which reaches the minimax optimal rate of convergence, such as $\overline{\theta}^{m_n^\star}$ in this example, is called minimax optimal. As in the oracle case, one should note that $m_n^\star$ in not available as it depends on $\Theta_{\mathfrak{a}}(r)$ and it is of interest to define a data-driven selection procedure for the parameter m which yields oracle and minimax optimality of the obtained estimator.

2.3 Model selection

A popular way to select the parameter m is the so-called model selection via penalised contrast. By properly choosing two functions $\mathrm{pen}:\mathbb{N}^\star\to\mathbb{R}_+$ and $\Upsilon:\mathbb{N}^\star\to\mathbb{R}_+$ respectively penalising the complexity of the chosen model (and hence increasing with m) and the loss of information by the cut-off (and hence decreasing with m), which do not depend on θ° nor $\Theta^{\mathfrak{a}}(r)$ but can depend on Y^n, one can adaptively select the cut-off parameter as $\widehat{m}:=\arg\min_{m\in[\![1,n]\!]}(\mathrm{pen}(m)+\Upsilon(m))$. We here justify by a Bayesian approach the choice $\mathrm{pen}(m)=3m/n$ and $\Upsilon(m)=-\sum_{j=1}^{m}Y_j^2$ and show in a novel way that it leads to minimax and oracle optimal estimation.

3 Bayesian inference

From a Bayesian point of view, one would define a prior distribution $\mathbb{P}^n_{\boldsymbol{\theta}}$ over the parameter space, potentially depending on the noise level n, after observing Y^n with likelihood $\mathbb{P}^n_{Y^n|\boldsymbol{\theta}}$ one would update the distribution of $\boldsymbol{\theta}$ to obtain the posterior distribution $\mathbb{P}^n_{\boldsymbol{\theta}|Y^n}$.
To analyse the quality of a Bayesian procedure from a frequentist point of view one needs to admit the existence of a true parameter θ° and proof that the posterior distribution contracts with optimal rate for some criterion.

3.1 Oracle optimality

An intuitive prior for a single element $\boldsymbol{\theta}_j$, j in $\mathbb{N}^\star$, of the sequence $\boldsymbol{\theta}$ would be a standard normal distribution $\mathbb{P}_{\boldsymbol{\theta}_j} = \mathcal{N}(0,1)$ as it is conjugate in this framework. Indeed, in our context, if we define, for any j in $\mathbb{N}^\star$,

$$\widehat{\theta}_j := \frac{n \cdot Y_j \cdot \lambda_j}{1 + n\lambda_j^2}; \quad \sigma_j := \frac{1}{1 + n\lambda_j^2};$$

we obtain $\mathbb{P}^n_{\boldsymbol{\theta}_j|Y^n} = \mathbb{P}^n_{\boldsymbol{\theta}_j|Y^n_j} = \mathcal{N}(\widehat{\theta}_j, \sigma_j)$.
We can then define the family of sieve priors, indexed by a parameter m in $\mathbb{N}^\star$ as the family of distributions $\mathcal{G} := \left\{ \mathbb{P}_{\boldsymbol{\theta}^m} = \bigotimes_{j=1}^{m} \mathcal{N}(0,1) \bigotimes_{j>m} \delta_0, m \in \mathbb{N}^\star \right\}$. The posterior then obtained is the Gaussian process with mean $\widehat{\theta}^m = \left(\widehat{\theta}_j \cdot \mathbb{1}_{j\leq m}\right)_{j\in\mathbb{N}}$ and variance $\sigma^m = (\sigma_j \cdot \mathbb{1}_{j\leq m})_{j\in\mathbb{N}}$.
One of the main result of this paper is showing a lower bound for the contraction rate of the priors of this family and exhibiting a prior of this family which reaches this bound without a log-loss term, therefore giving purely Bayesian formulation of oracle optimality.

THEOREM 3.1 For any θ° in Θ and increasing, unbounded sequence c_n, we have, with Φ°_n and m°_n as in DEFINITION 1.1

$$\begin{aligned} \lim_{n\to\infty} \sup_{\mathbb{P}_{\boldsymbol{\theta}^{m_n}} \in \mathcal{G}} \mathbb{E}^n_{\theta^\circ} \left[\mathbb{P}^n_{\boldsymbol{\theta}^{m_n}|Y^n} \left(d^2\left(\theta^\circ, \boldsymbol{\theta}\right) \leq c_n^{-1} \Phi^\circ_n \right) \right] &< 1; \\ \lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ} \left[\mathbb{P}^n_{\boldsymbol{\theta}^{m^\circ_n}|Y^n} \left(d^2\left(\theta^\circ, \boldsymbol{\theta}\right) \leq c_n \Phi^\circ_n \right) \right] &= 1. \end{aligned}$$

We therefore call Φ°_n oracle optimal contraction rate for the family $\mathcal{G}$ at θ°.

This result allows a purely Bayesian formulation of oracle optimality in the sense that it does not rely on comparison of the contraction rate with some convergence rate but with the contraction rates over a family of priors. Moreover, it shows contraction at the same rate as a popular family of estimators without a log-loss term in the upper bound.
However, the prior leading to optimal contraction rate depends on m°_n which is not available.

3.2 Minimax optimality

We also give attention here to Bayesian formulation of minimax optimality. The second major result of this paper shows that $\Psi^\star_n$ is a lower bound for the uniform contraction rate of posterior distributions and exhibit a posterior distribution reaching this rate without a log-loss term, giving purely Bayesian formulation of minimax optimality.

THEOREM 3.2 For any increasing and unbounded sequence, we have, with $\Psi_n^\star$ and $m_n^\star$ as in DEFINITION 1.1,

$$\lim_{n\to\infty} \sup_{\mathbb{Q}_{\boldsymbol{\theta}}} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}(r)} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{Q}^n_{\boldsymbol{\theta}|Y^n}\left(d^2\left(\theta^\circ,\boldsymbol{\theta}\right)\leq c_n^{-1}\cdot\Psi_n^\star(\Theta_{\mathfrak{a}}(r))\right)\right] < 1;$$
$$\lim_{n\to\infty} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}(r)} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\boldsymbol{\theta}^{m_n^\star}|Y^n}\left(d^2\left(\theta^\circ,\boldsymbol{\theta}\right)\leq c_n\cdot\Psi_n^\star(\Theta_{\mathfrak{a}}(r))\right)\right] = 1;$$

where $\sup_{\mathbb{Q}_{\boldsymbol{\theta}}}$ is taken over all prior distributions such that, for any positive sequence ρ_n, $\arg\max_{\widetilde{\theta}\in\Theta}\left\{\mathbb{Q}_{\boldsymbol{\theta}|Y^n}\left(\boldsymbol{\theta}\in\left\{\theta\in\Theta:\|\theta-\widetilde{\theta}\|^2\leq\rho_n\right\}\right)\right\}$ is an estimator.

We therefore call $\Psi_n^\star$ minimax optimal contraction over $\Theta^{\mathfrak{a}}(r)$ and $\mathbb{P}_{\boldsymbol{\theta}^{m_n^\star}}$ minimax optimal prior. Obviously $m_n^\star$ depends on $\Theta^{\mathfrak{a}}(r)$ and is not available in practice.
An important result given in Ghosal and van der Vaart already stated that a lower bound for the minimax optimal contraction rate is given by the minimax optimal convergence rate. It is however formulated for Hellinger distance and we give this result here for the L^2 distance, however, the proof follows the same lines.
A limitation of most results obtained about minimax contraction rates, up to our knowledge, is that the upper bounds differ from the lower bound by a log factor due to the fact that they derive from the powerful result obtained in Ghosal and van der Vaart which states general sufficient conditions on a prior sequence and parameter space to obtain a contraction rate. Moreover, most of those results are obtained for the Hellinger distance, which yields a weaker topology than the L^2 norm. However Nickl highlight results for the family of L^p-norms, including for $p=\infty$.

3.3 Hierarchical prior

From a Bayesian approach, a sensitive way to overcome the difficulty of selecting m is to use a so-called hierarchical prior where m is considered as a random variable M taking values in (a subset of) $\mathbb{N}^\star$. The prior on $\boldsymbol{\theta}$ is then denoted $\mathbb{P}_{\boldsymbol{\theta}^M}$ and is such that, for any m we have $\mathbb{P}_{\boldsymbol{\theta}^M|M=m}=\mathbb{P}_{\boldsymbol{\theta}^m}$. Moreover $Y^n,\boldsymbol{\theta}$ and M are such that $\mathbb{P}_{Y^n|\boldsymbol{\theta},M}=\mathbb{P}_{Y^n|\boldsymbol{\theta}}$. As a consequence, we have $\mathbb{P}_{\boldsymbol{\theta}^M|Y}=\sum_{m\in\mathbb{N}}\mathbb{P}_{\boldsymbol{\theta}^m|Y^n}\mathbb{P}_{M|Y^n}(m)$ and $\mathbb{P}_{M|Y^n}(m)=\frac{\int_\Theta \mathbb{P}_{Y^n|\boldsymbol{\theta}}\cdot\mathbb{P}_{\boldsymbol{\theta}^m}\cdot\mathbb{P}_M}{\mathbb{P}_{Y^n}}$.

In our case, following the methodology presented in JOHANNES ET AL. (2016), we define $G_n:=\max\{m\in[\![1,n]\!]:\Lambda_m/n\leq\Lambda_1\}$ and chose $\mathbb{P}^n_M(m)=\mathbb{1}_{m\in[\![1,G_n]\!]}\frac{\exp[-3m/2]}{\sum_{k=1}^{G_n}\exp[-3k/2]}$. This choice leads to the posterior distribution $\mathbb{P}^n_{M|Y^n}=\frac{\exp\left[-\frac{1}{2}\left(3m-\|\widehat{\theta}^m\|^2_{\sigma^m}\right)\right]}{\sum_{k=1}^{G_n}\exp\left[-\frac{1}{2}\left(3k-\|\widehat{\theta}^k\|^2_{\sigma^k}\right)\right]}$ and posterior mean $\widehat{\theta}^M=\sum_{m=1}^{G_n}\widehat{\theta}^m\mathbb{P}_{M|Y^n}(m)=\left(\widehat{\theta}_j\mathbb{P}_{M|Y^n}([\![j,G_n]\!])\right)_{j\in\mathbb{N}}$ where $\widehat{\theta}^m$, $\widehat{\theta}_j$ and σ^k are define in SECTION 3.1 and $\|\cdot\|_{\sigma^k}$ is defined as in SECTION 1.2 with the convention "$0/0=0$".

Interesting results about this posterior distribution and mean are already given in JOHANNES ET AL. (2016) they are reminded hereafter with a stronger, yet simpler to formulate, set of assumptions.

LEMMA 3.1 Under ASSUMPTION 1.1 and ASSUMPTION 1.2, 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 $K^\circ := 10(2 \vee \|\theta^\circ\|^2)L^2(16 \vee D^\circ \Lambda_{D^\circ})$ we have:

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

LEMMA 3.2 Under ASSUMPTION 1.1 and ASSUMPTION 1.3, if, in addition, $\log(G_n)/m_n^\star \to 0$ as $n \to \infty$ then

- for all θ° in $\Theta_{\mathfrak{a}}(r)$, with $D^\star := D^\star(\mathfrak{a}, \lambda) = \lceil 5L/\kappa^\star \rceil$ and $K^\star := 16\,(2 \vee r)\,L^2\,(16 \vee D^\star \Lambda_{D^\star})\,(1 \vee r)$, we have
$$\lim_{n\to\infty} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{P}_{\boldsymbol{\theta}^M|Y^n}^n \left(\|\theta^\circ - \boldsymbol{\theta}^M\|^2 \leq K^\star \Phi_n^\star\right)\right] = 1;$$
- for any monotonically increasing and unbounded sequence K_n holds
$$\lim_{n\to\infty} \inf_{\theta^\circ \in \Theta_{\mathfrak{a}}(r)} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{P}_{\boldsymbol{\theta}^M|Y^n}^n \left(\|\theta^\circ - \boldsymbol{\theta}^M\|^2 \leq K_n \Phi_n^\star\right)\right] = 1.$$

In the next section, we intend to generalise those results to a larger family of Bayesian methods.

4 Iterative procedure

An additional feature of this article is to provide a Bayesian interpretation of model selection through iteration of the hierarchical prior and to use this to propose a new proof of its optimality.

4.1 Principle

Consider an integer η, greater that 1, which is an iteration parameter. For $\eta = 1$, the situation is the regular Bayesian framework, with prior $\mathbb{P}_{\boldsymbol{\theta}}^n$, likelihood $\mathbb{P}_{Y^n|\boldsymbol{\theta}}^n$ and observations Y^n. For $\eta = 2$, the prior used, $\mathbb{P}_{\boldsymbol{\theta}}^{n,(2)}$ is the posterior for $\eta = 1$, that is to say $\mathbb{P}_{\boldsymbol{\theta}}^{n,(2)} = \mathbb{P}_{\boldsymbol{\theta}|Y^n}^n$, the likelihood and observation remain unchanged and the obtained posterior is denoted $\mathbb{P}_{\boldsymbol{\theta}|Y^n}^{n,(2)}$. For any value of η strictly greater than 1, the prior distribution is given by $\mathbb{P}_{\boldsymbol{\theta}}^{n,(\eta)} := \mathbb{P}_{\boldsymbol{\theta}|Y^n}^{n,(\eta-1)}$, the likelihood and observation are the same as previously.
This procedure gives more and more weight to the observations as the prior vanishes. The iteration parameter η could be interpreted as a measure of the trust given to the prior distribution.
A case of interest, studied in BUNKE AND JOHANNES (2005) in a general setting which however does not apply here, is when one lets the iteration parameter η tend to infinity. If the posterior distribution converges the limit is called self-informative Bayes carrier and denoted $\mathbb{P}_{\boldsymbol{\theta}|Y^n}^{n,(\infty)}$ and if this limit distribution admits a finite first moment, we call it self-informative limit.

We prove that the contraction rate of the posterior distribution for each fixed value of η remains unchanged, as well as the fact that the limit distribution is degenerated on a model selection estimator and provide a new proof of optimality of this estimator.

4.2 Iterated sieve prior

For any j and η in $\mathbb{N}^\star$, define $\widehat{\theta}_j^{(\eta)} := \frac{n\eta Y_j^n \lambda_j}{1+n\eta\lambda_j^2}$ and $\sigma_j^{(\eta)} := \frac{1}{1+n\eta\lambda_j^2}$. Then, for any m in $\mathbb{N}^\star$, if the prior chosen for $\eta = 1$ is $\mathbb{P}_{\boldsymbol{\theta}^m}$ as given in SECTION 3.1; after η iterations $(1 \leq \eta < \infty)$ the posterior obtained, $\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^m|Y^n}$, is the Gaussian process with mean $\widehat{\theta}^{m,(\eta)} := \left(\widehat{\theta}_j^{(\eta)} \mathbb{1}_{j\leq m}\right)_{j\in\mathbb{N}^\star}$ and variance $\sigma^{m,(\eta)} := \left(\sigma_j^{(\eta)} \mathbb{1}_{j\leq m}\right)_{j\in\mathbb{N}^\star}$.

Following the proofs of the results stated in SECTION 3.1 and SECTION 3.2 we derive the following results.

COROLLARY 4.1 For any θ° in Θ and increasing, unbounded sequence c_n, we have

$$\lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^{m_n^\circ}|Y^n}\left(d^2\left(\theta^\circ,\boldsymbol{\theta}\right) \leq c_n \Phi_n^\circ\right)\right] = 1.$$

COROLLARY 4.2 For any increasing and unbounded sequence, we have

$$\lim_{n\to\infty} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}(r)} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^{m_n^\star}|Y^n}\left(d^2\left(\theta^\circ,\boldsymbol{\theta}\right) \leq c_n \cdot \Psi_n^\star(\Theta_{\mathfrak{a}}(r))\right)\right] = 1;$$

Moreover, if one lets the number of iterations tend to infinity, we observe that the distribution degenerates around the projection estimator as defined in SECTION 2.1:

$$\lim_{\eta\to\infty} \mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^m|Y^n} = \delta_{\overline{\theta}^m}.$$

4.3 Iterated hierarchical prior

In the adaptive case, one needs to slightly modify the prior on M in order to conserve consistence as η tends to ∞. We hence define $\mathbb{P}^n_M(M=m) = \frac{\exp\left(-3\cdot\eta\cdot\frac{m}{2}\right)\cdot\prod_{j=1}^{m}\left(\frac{1}{\sigma_j^{(\eta)}}\right)^2}{\sum_{k=1}^{G_n}\exp\left(-3\cdot\eta\cdot\frac{k}{2}\right)\cdot\prod_{j=1}^{k}\left(\frac{1}{\sigma_j^{(\eta)}}\right)^2}$ with $\sigma_j^{(\eta)}$ as defined in SECTION 4.2.

Note that the prior depends on η, hence the interpretation of this parameter as an iteration parameter is a bit immoderate.

Hence, for all m in $[\![1, G_n]\!]$, the posterior distribution are characterised by :

$$\mathbb{P}^{n,(\eta)}_{M|Y^n}(m) = \frac{\exp\left[-\frac{1}{2}\left(3m\eta - \|\widehat{\theta}^{m,(\eta)}\|^2_{\sigma^{m,(\eta)}}\right)\right]}{\sum_{k=1}^{G_n}\exp\left[-\frac{1}{2}\left(3k\eta - \|\widehat{\theta}^{k,(\eta)}\|^2_{\sigma^{k,(\eta)}}\right)\right]},$$

and

$$\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^M|Y^n} = \sum_{m\in\mathbb{N}^\star}\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^m|Y^n}\cdot\mathbb{P}^{n,(\eta)}_{M|Y^n}(m);$$

and the posterior mean is then $\widehat{\theta}^{M,(\eta)} := \sum_{m\in\mathbb{N}^\star}\widehat{\theta}^{m,(\eta)}\mathbb{P}^{n,(\eta)}_{M|Y^n}(m) = \left(\widehat{\theta}_j^{(\eta)}\cdot\mathbb{P}^{n,(\eta)}_{M|Y^n}(M\geq j)\right)_{j\in\mathbb{N}^\star}$.

The posterior mean is both a shrinkage estimator as well as an aggregation estimator which aggregates optimally the posterior mean estimators of the priors of $\mathcal{G}$, giving weight $\mathbb{P}^{n,(\eta)}_{M|Y^n}(m)$ to the posterior mean of $\mathbb{P}_{\boldsymbol{\theta}^m}$.

As we have seen previously with the sieve priors, the iteration procedure conserves the contraction rate.

COROLLARY 4.3 Under ASSUMPTION 1.1 and ASSUMPTION 1.2, 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 $K^\circ := 10(2 \vee \|\theta^\circ\|^2) L^2 (16 \vee D^\circ \Lambda_{D^\circ})$ we have, for any η $(1 \leq \eta < \infty)$:

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

COROLLARY 4.4 Under ASSUMPTION 1.1 and ASSUMPTION 1.3, if, in addition, $\log(G_n)/m_n^\star \to 0$ as $n \to \infty$ then, for any η $(1 \leq \eta < \infty)$

- for all θ° in $\Theta_{\mathfrak{a}}(r)$, with $D^\star := D^\star(\mathfrak{a}, \lambda) = \lceil 5L/\kappa^\star \rceil$ and $K^\star := 16\,(2 \vee r)\, L^2\, (16 \vee D^\star \Lambda_{D^\star})\,(1 \vee r)$, we have
$$\lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ} \left[\mathbb{P}^{n,(\eta)}_{\boldsymbol{\theta}^M|Y^n} \left(\|\theta^\circ - \boldsymbol{\theta}^M\|^2 \leq K^\star \Phi_n^\star \right) \right] = 1;$$
- for any monotonically increasing and unbounded sequence K_n holds

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

Now, in this adaptive case, we consider the eventuality of letting η tend to infinity. In the spirit of the frequentist model selection method presented in SECTION 2.3, define $\Upsilon_\eta(m) = -\sum_{j=1}^{m} \frac{1}{1+\frac{\Lambda_j}{\eta n}} Y_j^2$ and $E_\eta(m) = \mathrm{pen}(m) + \Upsilon_\eta(m)$.
We see that for all m in $[\![1, G_n]\!]$,

$$\mathbb{P}^{n,(\eta)}_{M|Y^n}(m) = \frac{1}{\sum_{k=1}^{G_n} \exp\left[-\frac{\eta n}{2}\left(E_\eta(k) - E_\eta(m)\right)\right]}.$$

If η tends to $+\infty$, for all m, $\Upsilon_\eta(m)$ tends to $\Upsilon(m) := -\sum_{j=1}^{m} (Y_j)^2$ and we define for all m, $E(m) := \mathrm{pen}(m) + \Upsilon(m)$.

Interestingly, if we define the contrast Γ for any sequence $\theta^\star$ in Θ as

$$\Gamma\left(\theta^\star\right) := \sum_{j=1}^{G_n} \left(\theta_j^\star\right)^2 \lambda_j^2 - 2\sum_{j=1}^{G_n} \theta_j^\star \lambda_j Y_j,$$

we see, by differentiating Γ summand-wise, that $\overline{\theta}^{G_n}$ minimises this contrast and that $\Gamma\left(\overline{\theta}^{G_n}\right) = \Upsilon\left(G_n\right)$.

If for all k different from m, $E(k) - E(m) > 0$, then $\mathbb{P}^{n,(\eta)}_{M|Y^{(n)}}(m)$ trivially tends to 1 as η tends to ∞. On the other hand, if there exists k such that $E(k) - E(m) < 0$, then $\mathbb{P}^{n,(\eta)}_{M|Y^n}(m)$ obviously tends to 0 as η tends to ∞. So we see that, similarly to the model selection, this method only selects threshold parameters that minimise a penalised contrast.

Note that 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, $\mathbb{P}_{\theta^\circ}[\{\Upsilon(k) - \Upsilon(m) = \mathrm{pen}(k) - \mathrm{pen}(m)\}] = 0$.
We hence define $\widehat{m} := \underset{m \in [\![1, G_n]\!]}{\arg\min} \{E(m)\}$ and $\overline{\theta}^{\widehat{m}}$ the associated projection estimator. Hence, the self informative Bayes limit is $\overline{\theta}^{\widehat{m}}$ and the self informative Bayes carrier is degenerated on it: $\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M | Y^n} = \delta_{\overline{\theta}^{\widehat{m}}}$.

We obtain here optimality results both for the self informative limit and self informative Bayes carrier.

THEOREM 4.1 Consider $\overline{\theta}^{\widehat{m}}$ the frequentist estimator given by the self-informative limit. Under ASSUMPTION 1.1, ASSUMPTION 1.2 and the condition that $\limsup_{n \to \infty} \frac{\log\left(\frac{G_n^2}{\Phi_n^\circ}\right)}{m_n^\circ} \leq \frac{5}{9L}$, we have

$$\exists C^\circ \in \mathbb{R}_+^\star : \forall \theta^\circ \in \Theta, \quad \mathbb{E}^n_{\theta^\circ}\left[\|\overline{\theta}^{\widehat{m}} - \theta^\circ\|^2\right] \leq C^\circ \Phi_n^\circ.$$

This first theorem states that, under our set of assumptions, the self-informative limit reaches the oracle rate of the projection estimators.

THEOREM 4.2 Under ASSUMPTION 1.1, ASSUMPTION 1.2 and the condition that $\limsup_{\epsilon \to 0} \frac{\log(G_n)}{m_n^\circ}$, define $D^\circ := \left\lceil \frac{3}{\kappa^\circ} + 1 \right\rceil$ and $K^\circ := 16L \cdot [9 \vee D^\circ \Lambda_{D^\circ}]$; then, we have for all θ° in Θ,

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

This result states that the self informative Bayes carrier contracts with oracle optimal rate of the sieve priors under our set of assumptions.

THEOREM 4.3 Consider $\overline{\theta}^{\widehat{m}}$ the frequentist estimator given by the self-informative limit. Then, under ASSUMPTION 1.1, ASSUMPTION 1.3 and the condition that $\limsup_{n \to \infty} \frac{\log\left(\frac{G_n^2}{\Phi_n^\star}\right)}{m_n^\star} < \frac{5}{9L}$, we have

$$\exists C^\star \in \mathbb{R}_+^\star : \quad \sup_{\theta^\circ \in \Theta} \mathbb{E}^n_{\theta^\circ}\left[\|\overline{\theta}^{\widehat{m}} - \theta^\circ\|^2\right] \leq C^\star \Psi_n^\star.$$

This result shows that the self-informative limit converges with minimax optimal rate over Sobolev's ellipsoids under our set of assumptions.

THEOREM 4.4 Under ASSUMPTION 1.1, ASSUMPTION 1.3 and the condition that $\limsup_{n \to \infty} \frac{\log(G_n)}{m_n^\star}$, define $D^\star := \left\lceil \frac{3(1 \vee L^\circ)}{\kappa^\star} + 1 \right\rceil$ and $K^\star := 9L\,(1 \vee L^\circ)\, D^\star \Lambda_{D^\star}$; then, we have for all θ° in $\Theta^{\mathfrak{a}}(r)$,

$$\lim_{n \to \infty} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M | Y^n}\left(\|\boldsymbol{\theta}^M - \theta^\circ\|^2 \leq K^\star \Psi_n^\star\right)\right] = 1,$$

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

$$\lim_{n \to \infty} \sup_{\theta^\circ \in \Theta^{\mathfrak{a}}(r)} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M | Y^n}\left(\|\boldsymbol{\theta}^M - \theta^\circ\|^2 \leq K_n \Psi_n^\star\right)\right] = 1.$$

A Proof for THEOREM 3.1

A.1 Proof of the lower bound

We want to find a sequence $(K_n)_{n\in\mathbb{N}}$ (for short, K_n) converging to 0 such that

$$\lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] > 0.$$

For any n, we define $S^{m_n} := \sum_{j=1}^{m_n}(\boldsymbol{\theta}-\theta^\circ)^2$. Therefore we have :

$$\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] = \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(S^{m_n} \geq K_n - \mathfrak{b}_{m_n}\right)\right].$$

By definition, S^{m_n} has strictly positive, finite variance. We define $\mathcal{S}^{m_n} := \frac{S^{m_n}}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}|Y^n}[S^{m_n}]}}$. The sequence of random variables defined this way has variance 1 and is almost surely positive. Hence, for any ω, there exist a constant d strictly positive such that $\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\mathcal{S}^{m_n} \geq d\right) > 0$. Moreover we have :

$$\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] = \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\mathcal{S}^{m_n} \geq \frac{K_n - \mathfrak{b}^2_{m_n}}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right]}}\right)\right].$$

We now control the convergence in probability of $\mathbb{V}^n_{\boldsymbol{\theta}^m|Y^n}\left[\mathcal{S}^{m_n}\right]$ which is given by

$$\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right] = 2\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{n\eta}+1}\right)^2\left(1+2\frac{\left(-\theta^\circ_j+\eta\sqrt{n}\xi_j\lambda_j\right)^2}{\frac{\eta n}{\Lambda_j}\left(\frac{\Lambda_j}{\eta n}+1\right)}\right).$$

We now control the stochastic part of this object.
Define for some sequence $(u_n)_{n\in\mathbb{N}}$, tending to 0, the event $\Omega_{m_n} := \left\{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right] \geq u_n\right\}$. Which gives

$$\begin{aligned}\Omega_{m_n} &= \left\{2\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{n\eta}+1}\right)^2\left(1+2\frac{\left(-\theta^\circ_j+\eta\sqrt{n}\xi_j\lambda_j\right)^2}{\frac{\eta n}{\Lambda_j}\left(\frac{\Lambda_j}{\eta n}+1\right)}\right) \geq u_n\right\}\\ &= \left\{\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{n\eta}+1}\right)^3\left(-\theta^\circ_j+\eta\sqrt{n}\xi_j\lambda_j\right)^2 \geq \frac{u_n}{4} - \frac{1}{2}\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{n\eta}+1}\right)^2\right\}\end{aligned}$$

In the same spirit as previously, we define the sequence of random variables $T^{m_n} := \sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{n\eta}+1}\right)^3\left(-\theta^\circ_j+\eta\sqrt{n}\xi_j\lambda_j\right)^2$. We have

$$\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[T^{m_n}\right] = 2\sum_{j=1}^{m_n}\frac{1}{1+\frac{\Lambda_j}{\eta n}}\frac{\Lambda_j^4}{n^4\eta^2}\left[1+2\frac{\Lambda_j}{\eta^2 n}\left(\theta^\circ_j\right)^2\right]$$

and the sequence of random variables $\mathcal{T}^{m_n} := \frac{T^{m_n}}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[T^{m_n}]}}$ has variance 1 and is almost surely positive. Therefore, $\mathbb{P}^n_{\theta^\circ}(\Omega_{m_n}) = \mathbb{P}^n_{\theta^\circ}\left(\mathcal{T}^{m_n} \geq \frac{\frac{u_n}{4} - \frac{1}{2}\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{n\eta}+1}\right)^2}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[T^{m_n}]}}\right)$. Consider any strictly positive constant c such that $\lim_{n\to\infty}\mathbb{P}^n_{\theta^\circ}(\mathcal{T}^{m_n} \geq c) > 0$. Then, if

$$
\begin{aligned}
u_n =& 4\cdot c\cdot\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[T^{m_n}\right]} + 2\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{\eta n}+1}\right)^2 \\
=& 4\cdot c\cdot\sqrt{2\sum_{j=1}^{m_n}\frac{1}{1+\frac{\Lambda_j}{\eta n}}\frac{\Lambda_j^4}{n^4\eta^2}\left[1+2\frac{\Lambda_j}{\eta^2 n}\left(\theta_j^\circ\right)^2\right]} + 2\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta}\cdot\frac{1}{\frac{\Lambda_j}{\eta n}+1}\right)^2,
\end{aligned}
$$

we obtain $\mathbb{P}^n_{\theta^\circ}(\Omega_{m_n}) \geq \mathbb{P}^n_{\theta^\circ}(\mathcal{T}^{m_n} \geq c) > 0$.

We can now conclude about the posterior contraction by defining

$$
\begin{aligned}
K_n :=& \mathfrak{b}^2_{m_n} + d\cdot\sqrt{u_n} \\
=& \mathcal{O}\left(\frac{m_n\overline{\Lambda}_{m_n}}{n\eta} \vee \mathfrak{b}^2_{m_n}\right)
\end{aligned}
$$

Indeed :

$$
\begin{aligned}
\mathbb{E}^n_{\theta^\circ}\left[\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] =& \mathbb{E}^n_{\theta^\circ}\left[\mathbb{1}_{\Omega_{m_n}}\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] \\
&+ \mathbb{E}^n_{\theta^\circ}\left[\mathbb{1}_{\Omega^c_{m_n}}\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] \\
\geq& \mathbb{E}^n_{\theta^\circ}\left[\mathbb{1}_{\{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[S^{m_n}]\geq u_n\}}\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\mathcal{S}^{m_n} \geq \frac{K_n - \mathfrak{b}^2_{m_n}}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right]}}\right)\right] \\
\geq& \mathbb{E}^n_{\theta^\circ}\left[\mathbb{1}_{\{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[S^{m_n}]\geq u_n\}}\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\mathcal{S}^{m_n} \geq \frac{K_n - \mathfrak{b}^2_{m_n}}{\sqrt{u_n}}\right)\right] \\
\geq& \mathbb{P}^n_{\theta^\circ}(\Omega_{m_n})\cdot\mathbb{E}^n_{\theta^\circ}\left[\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\mathcal{S}^{m_n} \geq \frac{K_n - \mathfrak{b}^2_{m_n}}{\sqrt{u_n}}\right)\right] \\
\geq& \mathbb{P}^n_{\theta^\circ}(\Omega_{m_n})\cdot\mathbb{E}^n_{\theta^\circ}\left[\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\mathcal{S}^{m_n} \geq d\right)\right] \\
>& 0
\end{aligned}
$$

Not sure about the last step: we have existance of d for any ω, is it enough to justify existence of d such that the expected probability is not 0? If not, would $d_n \to 0$ work?

A.2 Proof for the upper bound

We want to find a sequence $(K_n)_{n\in\mathbb{N}}$ (for short, K_n) converging to 0 such that

$$
\lim_{n\to\infty}\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \geq K_n\right)\right] = 0.
$$

For any n, we define $S^{m_n} := \sum_{j=1}^{m_n} (\boldsymbol{\theta} - \theta^\circ)^2$. Therefore we have :

$$\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \geq K_n\right)\right] = \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(S^{m_n} \geq K_n - \mathfrak{b}_{m_n}\right)\right].$$

By definition, S^{m_n} has finite expectation and strictly positive, finite variance. We define $\mathcal{S}^{m_n} := \frac{S^{m_n} - \mathbb{E}^n_{\boldsymbol{\theta}|Y^n}[S^{m_n}]}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}|Y^n}[S^{m_n}]}}$. The sequence of random variables defined this way is tight as their expectations are all equal to 0 and their variances to 1. We now have to look for a contraction rate for this new family of random variables as we have :

$$\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \geq K_n\right)\right] = \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\theta^{m_n}|Y^n}\left(\mathcal{S}^{m_n} \geq \frac{K_n - \mathfrak{b}^2_{m_n} - \mathbb{E}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right]}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right]}}\right)\right].$$

We now control the convergence in probability of $\mathbb{E}^n_{\boldsymbol{\theta}^m|Y^n}\left[\mathcal{S}^{m_n}\right]$ and $\mathbb{V}^n_{\boldsymbol{\theta}^m|Y^n}\left[\mathcal{S}^{m_n}\right]$ which are given by

$$\begin{aligned}
\mathbb{E}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right] =& \sum_{j=1}^{m_n} \frac{\Lambda_j}{n\eta} \cdot \left(\frac{1}{\frac{\Lambda_j}{n\eta} + 1}\right)\left(1 + \frac{\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2}{\frac{\eta n}{\Lambda_j}\left(\frac{\Lambda_j}{\eta n} + 1\right)}\right) \\
\leq& \sum_{j=1}^{m_n} \frac{\Lambda_j}{n\eta} + \sum_{j=1}^{m_n} \frac{\Lambda_j^2}{n^2\eta^2}\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2; \\
\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[S^{m_n}\right] =& 2\sum_{j=1}^{m_n} \left(\frac{\Lambda_j}{n\eta} \cdot \frac{1}{\frac{\Lambda_j}{n\eta} + 1}\right)^2 \left(1 + 2\frac{\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2}{\frac{\eta n}{\Lambda_j}\left(\frac{\Lambda_j}{\eta n} + 1\right)}\right) \\
\leq& 2\sum_{j=1}^{m_n} \frac{\Lambda_j^2}{n^2\eta^2} + 4\sum_{j=1}^{m_n} \frac{\Lambda_j^3}{n^3\eta^3}\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2.
\end{aligned}$$

We now control the stochastic parts of those moments.
Define for some sequence $(u_n)_{n\in\mathbb{N}}$, tending to 0 and any deterministic sequence (a_j) the event $\Omega_{m_n} := \left\{\sum_{j=1}^{m_n} a_j\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2 \leq u_n\right\}$.
Obviously, $\Omega^c_{m_n} = \left\{\sum_{j=1}^{m_n} a_j\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2 \geq u_n\right\}$ has probability

$$\mathbb{P}^n_{\theta^\circ}\left(\Omega^c_{m_n}\right) = \mathbb{P}^n_{\theta^\circ}\left(\sum_{j=1}^{m_n}\left(-\theta^\circ_j + \eta\sqrt{n}\xi_j\lambda_j\right)^2 \geq u_n\right).$$

In the same spirit as previously, we define the sequence of random variables $T^{m_n} :=$

$\sum_{j=1}^{m_n} a_j \left(-\theta_j^\circ + \eta\sqrt{n}\xi_j\lambda_j\right)^2$. We have

$$\begin{aligned}
\mathbb{E}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[T^{m_n}\right] =& \sum_{j=1}^{m_n} a_j \frac{\eta^2 n}{\Lambda_j}\left[1 + \frac{\Lambda_j}{n\eta^2}\left(\theta_j^\circ\right)^2\right] \\
\leq & \sum_{j=1}^{m_n} a_j \frac{\eta^2 n}{\Lambda_j}\left[1 + \frac{\Lambda_1}{\eta^2}\left(\theta_j^\circ\right)^2\right]; \\
\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[T^{m_n}\right] =& 2\sum_{j=1}^{m_n} a_j^2 \left(\frac{\eta^2 n}{\Lambda_j}\right)^2\left[1 + 4\frac{\Lambda_j}{\eta^2 n}\left(\theta_j^\circ\right)^2\right] \\
\leq & 2\sum_{j=1}^{m_n} a_j^2 \left(\frac{\eta^2 n}{\Lambda_j}\right)^2\left[1 + 4\frac{\Lambda_1}{\eta^2}\left(\theta_j^\circ\right)^2\right];
\end{aligned}$$

and the sequence of random variables $\mathcal{T}^{m_n} := \frac{T^{m_n} - \mathbb{E}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[T^{m_n}]}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[T^{m_n}]}}$ is tight. Therefore, $\mathbb{P}^n_{\theta^\circ}\left(\sum_{j=1}^{m_n}\left(-\theta_j^\circ + \eta\sqrt{n}\xi_j\lambda_j\right)^2 \geq u_n\right) = \mathbb{P}^n_{\theta^\circ}\left(\mathcal{T}^{m_n} \geq \frac{u_n - \mathbb{E}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[T^{m_n}]}{\sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}[T^{m_n}]}}\right)$. Consider any sequence (c_n) diverging to infinity. Then if

$$\begin{aligned}
u_n =& \sqrt{\mathbb{V}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[T^{m_n}\right]}c_n + \mathbb{E}^n_{\boldsymbol{\theta}^{m_n}|Y^n}\left[T^{m_n}\right] \\
=& c_n \cdot \sqrt{2\sum_{j=1}^{m_n} a_j^2 \left(\frac{\eta^2 n}{\Lambda_j}\right)^2\left[1 + 4\frac{\Lambda_j}{\eta^2 n}\left(\theta_j^\circ\right)^2\right]} + \sum_{j=1}^{m_n} a_j \frac{\eta^2 n}{\Lambda_j}\left[1 + \frac{\Lambda_j}{n\eta^2}\left(\theta_j^\circ\right)^2\right].
\end{aligned}$$

Then $\mathbb{P}^n_{\theta^\circ}(\Omega^c_{m_n}) \leq \mathbb{P}^n_{\theta^\circ}\left(\mathcal{T}^{m_n} \geq c_n\right) \to 0$ as $\mathcal{T}^{m_n}$ is tight.

We can now conclude about the posterior contraction by defining

$$\begin{aligned}
K_n :=& \mathfrak{b}^2_{m_n} + \sum_{j=1}^{m_n} \frac{\Lambda_j}{n\eta} \cdot \left(\frac{1}{\frac{\Lambda_j}{n\eta} + 1}\right)\left(1 + \frac{u_n}{\frac{\eta n}{\Lambda_j}\left(\frac{\Lambda_j}{\eta n} + 1\right)}\right) \\
&+ c_n \cdot \sqrt{2\sum_{j=1}^{m_n}\left(\frac{\Lambda_j}{n\eta} \cdot \frac{1}{\frac{\Lambda_j}{n\eta} + 1}\right)^2 \left\{1 + 2\frac{u_n}{\frac{\eta n}{\Lambda_j}\left(\frac{\Lambda_j}{\eta n} + 1\right)}\right\}} \\
=& \mathcal{O}\left(c_n \cdot \frac{m_n \overline{\Lambda}_{m_n}}{n\eta} \vee \mathfrak{b}^2_{m_n}\right)
\end{aligned}$$

Indeed :

$$\begin{aligned}
\mathbb{E}^n_{\theta^\circ}\left[\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \geq K_n\right)\right] \leq & \mathbb{E}^n_{\theta^\circ}\left[\mathbb{1}_{\Omega n}\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \geq K_n\right)\right] + \mathbb{P}_{\theta^\circ}(\Omega n^c) \\
\leq & \mathbb{E}^n_{\theta^\circ}\left[\mathbb{Q}^n_{\theta^{m^\circ}|Y^n}\left(\mathcal{S}^{m_n} \geq c_n\right)\right] + \mathbb{P}_{\theta^\circ}(\Omega^c_n)
\end{aligned}$$

Which tends to 0 as $\mathcal{S}^{m_n}$ is a tight sequence of random variables. One could notice that if η diverges to infinity, the sequence c_n cancels and we recover the frequentist L_2 rate of convergence.

B Proof of THEOREM 3.2

B.1 Proof for the lower bound

We have already shown that, for any prior $\mathbb{P}_{\boldsymbol{\theta}}$, there exist an estimator $\widehat{\theta}$ such that, for any θ° in Θ and sequence Φ_n,

$$\lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \leq \Phi_n\right)\right] = 1 \Rightarrow \lim_{n\to\infty} \mathbb{P}^n_{\theta^\circ}\left(\|\widehat{\theta}-\theta^\circ\| \leq \Phi_n\right) = 1.$$

This implies that for any prior $\mathbb{P}_{\boldsymbol{\theta}}$ there exist an estimator $\widehat{\theta}$ such that, for any sequences $\mathfrak{a}_j$ and Ψ_n and real number r,

$$\lim_{n\to\infty} \inf_{\theta^\circ\in\Theta^{\mathfrak{a}}(r)} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \leq \Psi_n\right)\right] = 1 \Rightarrow \lim_{n\to\infty} \inf_{\theta^\circ\in\Theta^{\mathfrak{a}}(r)} \mathbb{P}^n_{\theta^\circ}\left(\|\widehat{\theta}-\theta^\circ\| \leq \Psi_n\right) = 1.$$

Which leads to the conclusion

$$\lim_{n\to\infty} \sup_{\mathbb{P}_{\boldsymbol{\theta}}} \inf_{\theta^\circ\in\Theta^{\mathfrak{a}}(r)} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}_{\boldsymbol{\theta}|Y^n}\left(\|\boldsymbol{\theta}-\theta^\circ\|^2 \leq \Psi_n\right)\right] = 1 \Rightarrow \lim_{n\to\infty} \sup_{\widehat{\theta}} \inf_{\theta^\circ\in\Theta^{\mathfrak{a}}(r)} \mathbb{P}^n_{\theta^\circ}\left(\|\widehat{\theta}-\theta^\circ\| \leq \Psi_n\right) = 1.$$

B.2 Proof for the upper bound

C Proof for THEOREM 4.1

Central results to proof the remaining theorems are the following.

LEMMA C.1 Let $\{X_j\}_{j\geq 1}$ be independent and normally distributed random variables with real mean α_j and standard deviation $\beta_j \geq 0$. For $m \in \mathbb{N}$, set $S_m := \sum_{j=1}^m X_j^2$ and consider $v_m \geq \sum_{j=1}^m \beta_j^2, t_m \geq \max_{1\leq j\leq m} \beta_j^2$ and $r_m \geq \sum_{j=1}^m \alpha_j^2$. Then for all $c \geq 0$, we have

$$\begin{aligned}
\sup_{m\geq 1} \exp\left[\frac{c(c\wedge 1)(v_m+2r_m)}{4t_m}\right] \mathbb{P}\left(S_m - \mathbb{E}[S_m] \leq -c(v_m+2r_m)\right) &\leq 1;\\
\sup_{m\geq 1} \exp\left[\frac{c(c\wedge 1)(v_m+2r_m)}{4t_m}\right] \mathbb{P}\left(S_m - \mathbb{E}[S_m] \geq \frac{3c}{2}(v_m+2r_m)\right) &\leq 1.
\end{aligned}$$

LEMMA C.2 Let $\{X_j\}_{j\geq 1}$ be independent and normally distributed random variables with real mean α_j and standard deviation $\beta_j \geq 0$. For $m \in \mathbb{N}$, set $S_m := \sum_{j=1}^m X_j^2$ and consider $v_m \geq \sum_{j=1}^m \beta_j^2, t_m \geq \max_{1\leq j\leq m} \beta_j^2$ and $r_m \geq \sum_{j=1}^m \alpha_j^2$. Then for all $c \geq 0$, we have

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

with $(a)_+ := (a \vee 0)$.

This proof being more complex, we will split it into several parts.

DEFINITION C.1 Define the following quantities :

$$G_n^- := \min \{ m \in [\![1, m_n^\circ]\!] : \quad \mathfrak{b}_m \leq 9L\Phi_n^\circ \},$$
$$G_n^+ := \max \left\{ m \in [\![m_n^\circ, G_n]\!] : \frac{(m - m_n^\circ)}{n} \leq 3\Lambda_{m_n^\circ}^{-1} \Phi_n^\circ \right\}.$$

PROPOSITION C.1 Under ASSUMPTION 1.1, we have the following concentration inequalities for the threshold hyper parameter :

$$\mathbb{P}_{\theta^\circ}^n \left[M > G_n^+ \right] \leq \exp\left[-\frac{5m_n^\circ}{9L} + \log(G_n) \right],$$
$$\mathbb{P}_{\theta^\circ}^n \left[M < G_n^- \right] \leq \exp\left[-\frac{7m_n^\circ}{9} + \log(G_n) \right].$$

C.1 Proof for PROPOSITION C.1

First, let's proof the first inequality. Use the fact that :

$$\begin{aligned}
\mathbb{P}_{\theta^\circ}\left[G_n^+ < \widehat{m} \leq G_n\right] =& \mathbb{P}_{\theta^\circ}\left[\forall l \in [\![1, G_n^+]\!], \quad \frac{3\widehat{m}}{n} - \sum_{j=1}^{\widehat{m}} Y_j^2 < \frac{3l}{n} - \sum_{j=1}^{l} Y_j^2\right] \\
\leq& \mathbb{P}_{\theta^\circ}\left[\exists m \in [\![G_n^+ + 1, G_n]\!] : \quad \frac{3m}{n} - \sum_{j=1}^{m} Y_j^2 < \frac{3m_n^\circ}{n} - \sum_{j=1}^{m_n^\circ} Y_j^2\right] \\
\leq& \sum_{m=G_n^+ + 1}^{G_n} \mathbb{P}_{\theta^\circ}\left[\frac{3m}{n} - \sum_{j=1}^{m} Y_j^2 < \frac{3m_n^\circ}{n} - \sum_{j=1}^{m_n^\circ} Y_j^2\right] \\
\leq& \sum_{m=G_n^+ + 1}^{G_n} \mathbb{P}_{\theta^\circ}\left[0 < \frac{3(m_n^\circ - m)}{n} + \sum_{j=m_n^\circ + 1}^{m} Y_j^2\right]
\end{aligned}$$

We will now use LEMMA C.1. For this purpose, define then for all m in $[\![G_n^+ + 1, G_n]\!]$: $S_m := \sum_{j=m_n^\circ+1}^{m} Y_j^2$, we then have $\mu_m := \mathbb{E}_{\theta^\circ}^n [S_m] = \frac{m - m_n^\circ}{n} + \sum_{j=m_n^\circ+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2$, $\alpha_j^2 := \mathbb{E}_{\theta^\circ}^n [Y_j]^2 = \left(\theta_j^\circ \lambda_j\right)^2$ and $\beta_j^2 := \mathbb{V}_{\theta^\circ}^n [Y_j] = \frac{1}{n}$.

Now, using that λ is monotonically decreasing and $\mathfrak{b}_{m_n^\circ} \leq \Phi_n^\circ$, we note

$$
\begin{aligned}
\sum_{j=m_n^\circ+1}^{m} \alpha_j^2 &= \sum_{j=m_n^\circ+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2 \\
&\leq \Lambda_{m_n^\circ}^{-1} \sum_{j=m_n^\circ+1}^{m} \left(\theta_j^\circ\right)^2 \\
&\leq \Lambda_{m_n^\circ}^{-1} \mathfrak{b}_{m_n^\circ} \\
&\leq \Lambda_{m_n^\circ}^{-1} \Phi_n^\circ && =: r_m \\
\sum_{j=m_n^\circ}^{m} \beta_j^2 &= \frac{m - m_n^\circ}{n} && =: v_m \\
\max_{j \in [\![m_n^\circ, m]\!]} \beta_j &= \frac{1}{n} && =: t_m
\end{aligned}
$$

Hence, we have, for all m in $[\![G_n^+, G_n]\!]$

$$
\begin{aligned}
\mathbb{P}_{\theta^\circ}^n \left[\sum_{j=m_n^\circ+1}^{m} Y_j^2 - 3\frac{m - m_n^\circ}{n} > 0 \right] &= \mathbb{P}_{\theta^\circ}^n \left[S_m - \frac{m - m_n^\circ}{n} > 2\frac{m - m_n^\circ}{n} \right] \\
&= \mathbb{P}_{\theta^\circ}^n \left[S_m - \frac{m - m_n^\circ}{n} - \sum_{j=m_n^\circ+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2 > 2\frac{m - m_n^\circ}{n} - \sum_{j=m_n^\circ+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2 \right] \\
&\leq \mathbb{P}_{\theta^\circ}^n \left[S_m - \mu_m > 2\frac{m - m_n^\circ}{n} - \Lambda_{m_n^\circ}^{-1} \mathfrak{b}_{m_n^\circ} \right].
\end{aligned}
$$

Using the definition of G_n^+, we have $\frac{m - m_n^\circ}{n} > 3\Lambda_{m_n^\circ}^{-1} \Phi_n^\circ$.
Hence, we can write, using ASSUMPTION 1.1 and LEMMA C.1 with $c = 2/3$:

$$
\begin{aligned}
\mathbb{P}_{\theta^\circ}^n \left[\sum_{j=m_n^\circ}^{m} Y_j^2 - 3\frac{m - m_n^\circ}{n} > 0 \right] &\leq \mathbb{P}_{\theta^\circ}^n \left[S_m - \mu_m > \frac{m - m_n^\circ}{n} + 2\Lambda_{m_n^\circ}^{-1} \Phi_n^\circ \right] \\
&\leq \mathbb{P}_{\theta^\circ}^n \left[S_m - \mu_m > v_m + 2r_m \right] \\
&\leq \exp\left[-n \frac{\frac{m - m_n^\circ}{n} + 2\Lambda_{m_n^\circ}^{-1} \Phi_n^\circ}{9} \right] \\
&\leq \exp\left[-n \frac{5\Lambda_{m_n^\circ}^{-1} \Phi_n^\circ}{9} \right] \\
&\leq \exp\left[-\frac{5 m_n^\circ}{9L} \right].
\end{aligned}
$$

Finally we can conclude that

$$
\mathbb{P}_{\theta^\circ}^n \left[G_n^+ < \widehat{m} \leq G_n \right] \leq \exp\left[-\frac{5 m_n^\circ}{9L} + \log\left(G_n\right) \right].
$$

We now prove the second inequality.
We begin by writing the same kind of inclusion of events as for the first inequality :

$$\begin{aligned}
\mathbb{P}^n_{\theta^\circ}\left[1 \leq \widehat{m} < G_n^-\right] =& \mathbb{P}^n_{\theta^\circ}\left[\forall m \in [\![G_n^-, G_n]\!], \quad \Upsilon(\widehat{m}) + \mathrm{pen}(\widehat{m}) < \Upsilon(m) + \mathrm{pen}(m)\right] \\
\leq & \mathbb{P}_{\theta^\circ}\left[\exists m \in [\![1, G_n^- - 1]\!], \quad \Upsilon(m) + \mathrm{pen}(m) < \gamma(m_n^\circ) + \mathrm{pen}(m_n^\circ)\right] \\
\leq & \sum_{m=1}^{G_n^-} \mathbb{P}^n_{\theta^\circ}\left[\Upsilon(m) + \mathrm{pen}(m) < \Upsilon(m_n^\circ) + \mathrm{pen}(m_n^\circ)\right] \\
\leq & \sum_{m=1}^{G_n^-} \mathbb{P}^n_{\theta^\circ}\left[\sum_{j=m+1}^{m_n^\circ} Y_j^2 < 3\frac{m_n^\circ - m}{n}\right].
\end{aligned}$$

The LEMMA C.1 steps in again. Define $S_m := \sum_{j=m+1}^{m_n^\circ} Y_j^2$ and we want to control the concentration of this sum, hence we take the following notations :

$$\begin{aligned}
\mu_m :=& \mathbb{E}_{\theta^\circ}[S_m] \\
=& \frac{m_n^\circ - m}{n} + \sum_{j=m+1}^{m_n^\circ} \left(\theta_j^\circ \lambda_j\right)^2 \\
r_m :=& \sum_{j=m+1}^{m_n^\circ} \left(\theta_j^\circ \lambda_j\right)^2 \\
v_m :=& \frac{m_n^\circ - m}{n} \\
t_m :=& \frac{1}{n}.
\end{aligned}$$

Hence, we have, using ASSUMPTION 1.1 Constant last line: obtained $3L$ now obtain $5L$

$$\begin{aligned}
\mathbb{P}^n_{\theta^\circ}\left[S_m < 3\frac{m_n^\circ - m}{n}\right] =& \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < 3\frac{m_n^\circ - m}{n} - \frac{m_n^\circ - m}{n} - \sum_{j=m+1}^{m_n^\circ} \left(\theta_j^\circ \lambda_j\right)^2\right] \\
=& \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < 3\frac{m_n^\circ - m}{n} - \frac{2}{3}\frac{m_n^\circ - m}{n} - \frac{1}{3}\sum_{j=m+1}^{m_n^\circ} \left(\theta_j^\circ \lambda_j\right)^2 - \frac{1}{3}\left[v_m + 2r_m\right]\right] \\
\leq & \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < 3\frac{m_n^\circ - m}{n} - \frac{2}{3}\frac{m_n^\circ - m}{n} - \frac{1}{3}\Lambda^{-1}_{m_n^\circ}\sum_{j=m+1}^{m_n^\circ} \left(\theta_j^\circ\right)^2 - \frac{1}{3}\left[v_m + 2r_m\right]\right] \\
\leq & \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < -\frac{1}{3}\left[v_m + 2r_m\right] + \frac{7}{3}\frac{m_n^\circ}{n} + \frac{1}{3}\Lambda^{-1}_{m_n^\circ}\mathfrak{b}_{m_n^\circ} - \frac{1}{3}\Lambda^{-1}_{m_n^\circ}\mathfrak{b}_m\right] \\
\leq & \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < -\frac{1}{3}\left[v_m + 2r_m\right] + 3L\Phi_n^\circ\Lambda^{-1}_{m_n^\circ} - \frac{1}{3}\Lambda^{-1}_{m_n^\circ}\mathfrak{b}_m\right]
\end{aligned}$$

now, using the definition of G_n^-, we have $\mathfrak{b}_m > 9L\Phi_n^\circ$ so, using LEMMA C.1 after checking constant: $9L$ or $15L$

$$
\begin{aligned}
\mathbb{P}^n_{\theta^\circ}\left[S_m < 3\frac{m^\circ_n - m}{n}\right] \leq& \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < -\frac{1}{3}\left[v_m + 2r_m\right]\right] \\
\leq& \exp\left[-n\frac{\frac{m^\circ_n - m}{n} + 2\sum_{j=m+1}^{m^\circ_n}\left(\theta^\circ_j\lambda_j\right)^2}{36}\right] \\
\leq& \exp\left[-n\frac{\frac{m^\circ_n - m}{n} + 2\Lambda^{-1}_{m^\circ_n}\mathfrak{b}_m - 2\Lambda^{-1}_{m^\circ_n}\mathfrak{b}_{m^\circ_n}}{36}\right] \\
\leq& \exp\left[-n\frac{16L\Phi^\circ_n\Lambda^{-1}_{m^\circ_n}}{36}\right] \\
\leq& \exp\left[-\frac{4m^\circ_n}{9}\right]
\end{aligned}
$$

after checking constant: $16L$ or $28L$ => $4/9$ or $7/9$

Which in turn implies

$$
\mathbb{P}^\circ_{\theta^\circ}\left[1 \leq \widehat{m} < G^-_n\right] \leq \exp\left[-\frac{4m^\circ_n}{9} + \log\left(G_n\right)\right].
$$

C.2 Proof of the the final statement

The L^2 risk can be written :

$$
\mathbb{E}^n_{\theta^\circ}\left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] = \mathbb{E}^n_{\theta^\circ}\left[\sum_{j=1}^{G_n}\left(\overline{\theta}^{\widehat{m}}_j - \theta^\circ_j\right)^2\right] + \mathbb{E}^n_{\theta^\circ}\left[\sum_{j=G_n+1}^{\infty}\left(\theta^\circ_j\right)^2\right].
$$

Together with

$$
\forall j \in [\![1, G_n]\!], \quad \overline{\theta}^{\widehat{m}}_j - \theta^\circ_j = \left(\frac{Y_j}{\lambda_j} - \theta^\circ_j\right)\mathbb{1}_{\{\widehat{m}\in[\![j,G_n]\!]\}} + \theta^\circ_j\mathbb{1}_{\{\widehat{m}\in[\![1,j-1]\!]\}},
$$

implies that

$$
\begin{aligned}
\mathbb{E}^n_{\theta^\circ}\left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] \leq& \underbrace{\sum_{j=1}^{G_n}\mathbb{E}^n_{\theta^\circ}\left[\left(\frac{Y_j}{\lambda_j} - \theta^\circ_j\right)^2\mathbb{1}_{\{\widehat{m}\in[\![j,G_n]\!]\}}\right]}_{=:A} \\
&+ \underbrace{\sum_{j=1}^{G_n}\mathbb{E}^n_{\theta^\circ}\left[\left(\theta^\circ_j\right)^2\mathbb{1}_{\{\widehat{m}\in[\![1,j-1]\!]\}}\right]}_{=:B} + \underbrace{\sum_{j=G_n+1}^{\infty}\mathbb{E}^n_{\theta^\circ}\left[\left(\theta^\circ_j\right)^2\right]}_{=:C}.
\end{aligned}
$$

We will now control each of the three parts of the sum using LEMMA C.2 and PROPOSITION C.1.

First, consider A and let be some positive real number p to be specified later.
Then we can write

$$
\begin{aligned}
A \leq & \sum_{j=1}^{G_n^+} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \sum_{j=G_n^+ + 1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \mathbb{1}_{\{\widehat{m} \in [\![j, G_n]\!]\}} \right] \\
\leq & \sum_{j=1}^{G_n^+} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \sum_{j=G_n^+ + 1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]\}} \right] \\
\leq & \sum_{j=1}^{G_n^+} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \sum_{j=G_n^+ + 1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]\}} \right] \\
& - p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]} \right] + p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]} \right] \\
\leq & \sum_{j=1}^{G_n^+} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \mathbb{E}_{\theta^\circ}^n \left[\left(\sum_{j=G_n^+ + 1}^{G_n} \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 - p \right) \mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]\}} \right] \\
& + p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]} \right] \\
\leq & \underbrace{\sum_{j=1}^{G_n^+} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right]}_{=:A_3} + \underbrace{\mathbb{E}_{\theta^\circ}^n \left[\left(\sum_{j=G_n^+ + 1}^{G_n} \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 - p \right)_+ \right]}_{=:A_1} + \underbrace{p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^+ + 1, G_n]\!]} \right]}_{=:A_2}.
\end{aligned}
$$

First, we will control A_1 using LEMMA C.2.
The goal is to give p a value that is large enough to control this object but small enough so $p \cdot \mathbb{P}_{\theta^\circ}^n \left[G_n^+ < \widehat{m} \leq G_n \right]$ is still for the most Φ_n°.
Define $S_n := \sum_{j=G_n^+ + 1}^{G_n} \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2$.
We have, for all j in $[\![G_n^+ + 1, G_n]\!]$,

$$
\frac{Y_j}{\lambda_j} - \theta_j^\circ \sim \mathcal{N} \left(0, \frac{\Lambda_j}{n} \right),
$$

so $\mathbb{E}_{\theta^\circ}^n \left[S_n \right] = \frac{1}{n} \sum_{j=G_n^+ + 1}^{G_n} \Lambda_j$.
And define, using the definition of G_n and ASSUMPTION 1.1

$$
\begin{aligned}
t_m := & \Lambda_1 \geq \frac{\Lambda_{G_n}}{n} \geq \max_{j \in [\![G_n^+ + 1, G_n]\!]} \frac{\Lambda_j}{n} \\
v_m := & G_n \Lambda_1 \geq \frac{G_n \Lambda_{G_n}}{n} \geq \frac{1}{n} \sum_{j=G_n^+ + 1}^{G_n} \Lambda_j.
\end{aligned}
$$

We can take $p = \mathbb{E}^n_{\theta^\circ}[S_n] + 3v_m$ which gives, using the definition of G^+_n and $G_n > G^+_n$

$$\begin{aligned}
A_1 =& \mathbb{E}^n_{\theta^\circ}\left[(S_n - \mathbb{E}^n_{\theta^\circ}[S_n] - 3v_m)_+\right] \\
\leq & 6\Lambda_1 \exp\left[-\frac{G_n}{2}\right] \\
\leq & 6\Lambda_1 \exp\left[-\frac{nG_n}{2n}\right] \\
\leq & 6\Lambda_1 \exp\left[-n\frac{3\Lambda^{-1}_{m^\circ_n}\Phi^\circ_n}{2} - \frac{m^\circ_n}{2}\right] \\
A_1 \leq & 6\Lambda_1 \exp\left[-\frac{2m^\circ_n}{L}\right].
\end{aligned}$$

Thanks to PROPOSITION C.1 it is easily shown that

$$A_2 < 4\Lambda_1 \exp\left[-\frac{5m^\circ_n}{9L} + 2\log(G_n)\right].$$

Finally, we control A_3. Using the definition of G^+_n we have

$$\begin{aligned}
A_3 =& \sum_{j=1}^{G^+_n} \mathbb{E}^n_{\theta^\circ}\left[\left(\overline{\theta}_j - \theta^\circ_j\right)^2\right] \\
=& \sum_{j=1}^{G^+_n} \frac{\Lambda_j}{n} \\
=& \frac{1}{n} G^+_n \overline{\Lambda}_{G^+_n} \\
A_3 \leq & \frac{\overline{\Lambda}_{G^+_n}}{\Lambda_{m^\circ_n}} 3\Phi^\circ_n.
\end{aligned}$$

Note that, using ASSUMPTION 1.2 and the definition of G^+_n, we have that $\frac{\overline{\Lambda}_{G^+_n}}{\Lambda_{m^\circ_n}}$ is bounded for n large enough; indeed with $D^\circ := \lceil \frac{3}{\kappa^\circ} + 1 \rceil$,

$$\begin{aligned}
G^+_n \leq \frac{3n\Phi^\circ_n}{\Lambda_{m^\circ_n}} + m^\circ_n \leq \frac{3nm^\circ_n \overline{\Lambda}_{m^\circ_n}}{\Lambda_{m^\circ_n} n\kappa^\circ} + m^\circ_n \leq \left(\frac{3}{\kappa^\circ} + 1\right) m^\circ_n \leq D^\circ m^\circ_n \\
\Rightarrow \frac{\overline{\Lambda}_{G^+_n}}{\Lambda_{m^\circ_n}} \leq \frac{\overline{\Lambda}_{D^\circ \cdot m^\circ_n}}{\Lambda_{D^\circ \cdot m^\circ_n}} \cdot \frac{\Lambda_{D^\circ \cdot m^\circ_n}}{\Lambda_{m^\circ_n}} \leq \Lambda_{D^\circ}.
\end{aligned}$$

Hence, we have

$$A \leq 6\Lambda_1 \exp\left[-\frac{2m^\circ_n}{L}\right] + 4\Lambda_1 \exp\left[-\frac{5m^\circ_n}{9L} + 2\log(G_n)\right] + \Lambda_{D^\circ} 3\Phi^\circ_n.$$

Now we control B We use a decomposition similar to the one used for A :

$$\begin{aligned}
B \leq & \sum_{j=G_n^-+1}^{G_n} \left(\theta_j^\circ\right)^2 + \sum_{j=1}^{G_n^-} \mathbb{E}_{\theta^\circ}^n \left[\left(\theta_j^\circ\right)^2 \mathbb{1}_{\{\widehat{m}\in[\![1,j-1]\!]\}}\right] \\
\leq & \sum_{j=G_n^-+1}^{G_n} \left(\theta_j^\circ\right)^2 + \sum_{j=1}^{G_n^-} \mathbb{E}_{\theta^\circ}^n \left[\left(\theta_j^\circ\right)^2 \mathbb{1}_{\{\widehat{m}\in[\![1,G_n^--1]\!]\}}\right] \\
B \leq & \underbrace{\sum_{j=G_n^-+1}^{G_n} \left(\theta_j^\circ\right)^2}_{=:B_1} + \underbrace{\sum_{j=1}^{G_n^-} \left(\theta_j^\circ\right)^2 \mathbb{P}_{\theta^\circ}^n \left[1 \leq \widehat{m} < G_n^-\right]}_{=:B_2}.
\end{aligned}$$

First, notice that $B_1 + C = \mathfrak{b}_{G_n^-} \leq 9L\Phi_n^\circ$ by the definition of G_n^-.

To control B_2, we use the fact that θ° is square summable and the PROPOSITION C.1 :

$$\begin{aligned}
B_2 = & \mathbb{P}_{\theta^\circ}^n \left[1 \leq \widehat{m} < G_n^-\right] \sum_{j=1}^{G_n^-} \left(\theta_j^\circ\right)^2 \\
\leq & \exp\left[-\frac{7m_n^\circ}{9} + \log\left(G_n\right)\right] \cdot \|\theta^\circ\|^2.
\end{aligned}$$

So we have

$$B + C \leq 9L\Phi_n^\circ + \|\theta^\circ\|^2 \cdot \exp\left[-\frac{7m_n^\circ}{9} + \log\left(G_n\right)\right].$$

Which leads to :

$$\begin{aligned}
\mathbb{E}_{\theta^\circ}^n \left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] \leq & 3\left(\Lambda_{D^\circ} + 3L\right)\Phi_n^\circ + \\
& \left(6\Lambda_1 \exp\left[-\frac{2m_n^\circ}{L} - \log\left(\Phi_n^\circ\right)\right] + 4\Lambda_1 \exp\left[-\frac{5m_n^\circ}{9L} + \log\left(\frac{G_n^2}{\Phi_n^\circ}\right)\right]\right. \\
& \left. + \|\theta^\circ\|^2 \cdot \exp\left[-\frac{7m_n^\circ}{9} + \log\left(\frac{G_n}{\Phi_n^\circ}\right)\right]\right)\Phi_n^\circ,
\end{aligned}$$

which proves that there exist C such that, for n large enough,

$$\mathbb{E}_{\theta^\circ}^n \left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] \leq C\Phi_n^\circ.$$

D Proof of THEOREM 4.2

To prove this theorem we need a stronger statement about the contraction rate for fixed m.

PROPOSITION D.1 Under ASSUMPTION 1.1, we have, for all m in $[\![1, G_n]\!]$

$$\mathbb{P}^n_{\theta^\circ}\left[\left\|\overline{\theta}^m - \theta^\circ\right\|^2 < \frac{1}{2}\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]\right] \leq \exp\left[-\frac{m}{16L}\right],$$
$$\mathbb{P}^n_{\theta^\circ}\left[\left\|\overline{\theta}^m - \theta^\circ\right\|^2 > 4\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]\right] \leq \exp\left[-\frac{m}{9L}\right].$$

D.1 Proof of PROPOSITION D.1

Let be m in $[\![1, G_n]\!]$ and note that

$$\|\overline{\theta}^m - \theta^\circ\|^2 = \sum_{j=1}^{m}\left(\frac{Y_j}{\lambda_j} - \theta^\circ_j\right)^2 + \mathfrak{b}_m.$$

We will use LEMMA C.1; therefor define

$$\begin{aligned}
& S_m := \sum_{j=1}^{m}\left(\frac{Y_j}{\lambda_j} - \theta^\circ_j\right)^2 &&, \\
& \mu_m := \mathbb{E}^n_{\theta^\circ}[S_m] && = \frac{m\overline{\Lambda}_m}{n}, \\
\forall j \in [\![1, m]\!] \quad & \beta_j^2 := \mathbb{V}^n_{\theta^\circ}\left[\frac{Y_j}{\lambda_j} - \theta^\circ_j\right] && = \frac{\Lambda_j}{n}, \\
\forall j \in [\![1, m]\!] \quad & \alpha_j^2 := \mathbb{E}^n_{\theta^\circ}\left[\frac{Y_j}{\lambda_j} - \theta^\circ_j\right] && = 0, \\
& v_m := \sum_{j=1}^{m}\beta_j^2 && = \frac{m\overline{\Lambda}_m}{n}, \\
& t_m := \max_{j\in[\![1,m]\!]}\beta_j^2 && = \frac{\Lambda_m}{n}.
\end{aligned}$$

We then control the concentration of S_m, first from above :

$$\begin{aligned}
\exp\left[-\frac{m}{16L}\right] &\geq \exp\left[-\frac{nm\overline{\Lambda}_m}{16n\Lambda_m}\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m \leq -\frac{1}{2}\frac{m\overline{\Lambda}_m}{n}\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m \leq \frac{1}{2}\frac{m\overline{\Lambda}_m}{n}\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \leq \frac{1}{2}\frac{m\overline{\Lambda}_m}{n} + \mathfrak{b}_m\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \leq \frac{1}{2}\left(\frac{m\overline{\Lambda}_m}{n} + \mathfrak{b}_m\right)\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \leq \frac{1}{2}\left[\frac{m\overline{\Lambda}_m}{n} \vee \mathfrak{b}_m\right]\right];
\end{aligned}$$

then from bellow :

$$
\begin{aligned}
\exp\left[-\frac{m}{9L}\right] \geq \exp\left[-\frac{nm\overline{\Lambda}_m}{9n\Lambda_m}\right]
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m \geq \frac{m\overline{\Lambda}_m}{n}\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m \geq 2\frac{m\overline{\Lambda}_m}{n}\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \geq 2\frac{m\overline{\Lambda}_m}{n} + \mathfrak{b}_m\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \geq 2\left(\frac{m\overline{\Lambda}_m}{n} + \mathfrak{b}_m\right)\right] \\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \geq 4\left[\frac{m\overline{\Lambda}_m}{n} \vee \mathfrak{b}_m\right]\right].
\end{aligned}
$$

D.2 Proof of the complete statement

By the the total probability formula, we have :

$$
\begin{aligned}
&\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left((K^\circ)^{-1}\Phi^\circ_n \leq \|\boldsymbol{\theta}^M - \theta^\circ\|^2 \leq K^\circ\Phi^\circ_n\right)\right] \\
&\quad = 1 - \underbrace{\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left((K^\circ)^{-1}\Phi^\circ_n > \|\boldsymbol{\theta}^M - \theta^\circ\|^2\right)\right]}_{=:A} - \underbrace{\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(K^\circ\Phi^\circ_n < \|\boldsymbol{\theta}^M - \theta^\circ\|^2\right)\right]}_{B}.
\end{aligned}
$$

Hence, we will control A and B separately.

We first control A :

$$
\begin{aligned}
A &= \sum_{m=1}^{G_n} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\left\{(K^\circ)^{-1}\Phi^\circ_n > \|\boldsymbol{\theta}^M - \theta^\circ\|^2\right\} \cap \{M = m\}\right)\right] \\
&\leq \sum_{m=1}^{G_n^- - 1} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{M|Y^n}(\{M = m\})\right] + \sum_{m=G_n^+ + 1}^{G_n} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{M|Y}(\{M = m\})\right] \\
&\quad + \sum_{m=G_n^-}^{G_n^+} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y,M=m}\left(\left\{(K^\circ)^{-1}\Phi^\circ_n > \|\boldsymbol{\theta}^M - \theta^\circ\|^2\right\}\right)\right] \\
&\leq \underbrace{\sum_{m=1}^{G_n^- - 1} \mathbb{P}^n_{\theta^\circ}[\{\widehat{m} = m\}]}_{=:A_1} + \underbrace{\sum_{m=G_n^+ + 1}^{G_n} \mathbb{P}^n_{\theta^\circ}[\{\widehat{m} = m\}]}_{=:A_2} + \sum_{m=G_n^-}^{G_n^+} \underbrace{\mathbb{P}^n_{\theta^\circ}\left[\left\{(K^\circ)^{-1}\Phi^\circ_n > \left\|\overline{\theta}^m - \theta^\circ\right\|^2\right\}\right]}_{=:A_{3,m}}.
\end{aligned}
$$

While A_1 and A_2 can respectively be controlled thanks to PROPOSITION C.1 by

$$
\begin{aligned}
A_1 &\leq \exp\left[-\frac{5m^\circ_n}{9L} + \log(G_n)\right] \\
A_2 &\leq \exp\left[-\frac{7m^\circ_n}{9} + \log(G_n)\right],
\end{aligned}
$$

we now have to control $\sum_{m=G_n^-}^{G_n^+} A_{3,m}$.

Thanks to PROPOSITION D.1, we have that, for all m in $[\![1, G_n]\!]$:

$$\mathbb{P}_{\theta^\circ}^n \left[\left\{\left\|\overline{\theta}^m - \theta^\circ\right\|^2 < \frac{1}{2}\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]\right\}\right] \leq \exp\left[-\frac{m}{16L}\right].$$

Moreover, by definition of Φ_n°, we have for all m in $[\![G_n^-, G_n^+]\!]$ that $\Phi_n^\circ \leq \left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]$, which implies, with $K^\circ \geq 16L$,

$$\begin{aligned}\mathbb{P}_{\theta^\circ}^n \left[\left\{\left\|\overline{\theta}^m - \theta^\circ\right\|^2 < (K^\circ)^{-1}\Phi_n^\circ\right\}\right] \leq & \mathbb{P}_{\theta^\circ}^n \left[\left\{\left\|\overline{\theta}^m - \theta^\circ\right\|^2 < \frac{1}{2}\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]\right\}\right] \\ \leq & \exp\left[-\frac{m}{K^\circ}\right].\end{aligned}$$

This allows us to conclude that

$$\begin{aligned}\sum_{m=G_n^-}^{G_n^+} A_{3,m} \leq & \sum_{m=G_n^-}^{G_n^+} \exp\left[-\frac{m}{K^\circ}\right] \\ \leq & \sum_{m=G_n^-}^{\infty} \exp\left[-\frac{m}{K^\circ}\right] \\ \leq & \int_{m=G_n^- -1}^{\infty} \exp\left[-\frac{m}{K^\circ}\right] dm \\ \leq & \left[K^\circ \exp\left[-\frac{m}{K^\circ}\right]\right]_{G_n^- -1}^{\infty} dm \\ \leq & K^\circ \exp\left[-\frac{G_n^- - 1}{K^\circ}\right]\end{aligned}$$

We now control B :

$$\begin{aligned}B = & \sum_{m=1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{P}_{\boldsymbol{\theta}^M|Y^n}^{n,(\infty)}\left(\{K^\circ\Phi_n^\circ < \|\boldsymbol{\theta}^M - \theta^\circ\|^2\} \cap \{M = m\}\right)\right] \\ \leq & \sum_{m=1}^{G_n^- -1} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{P}_{\boldsymbol{\theta}^M|Y^n}^{n,(\infty)}(\{M = m\})\right] + \sum_{m=G_n^+ +1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{P}_{M|Y^n}^{n,(\infty)}(\{M = m\})\right] \\ & + \sum_{m=G_n^-}^{G_n^+} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{P}_{\boldsymbol{\theta}^M|Y,M=m}^{n,(\infty)}\left(\{K^\circ\Phi_n^\circ < \|\boldsymbol{\theta}^M - \theta^\circ\|^2\}\right)\right] \\ \leq & \underbrace{\sum_{m=1}^{G_n^- -1} \mathbb{P}_{\theta^\circ}^n[\{\widehat{m} = m\}]}_{=:B_1} + \underbrace{\sum_{m=G_n^+ +1}^{G_n} \mathbb{P}_{\theta^\circ}^n[\{\widehat{m} = m\}]}_{=:B_2} + \sum_{m=G_n^-}^{G_n^+} \underbrace{\mathbb{P}_{\theta^\circ}^n \left[\left\{K^\circ\Phi_n^\circ < \left\|\overline{\theta}^m - \theta^\circ\right\|^2\right\}\right]}_{=:B_{3,m}}.\end{aligned}$$

As previously, B_1 and B_2 are controlled in PROPOSITION C.1. Hence, we now control

$$\sum_{m=G_n^-}^{G_n^+} B_{3,m}.$$

Using PROPOSITION D.1 again, we have that, for all m in $[\![1, G_n]\!]$:

$$\mathbb{P}_{\theta^\circ}^n \left[\left\{ \|\overline{\theta}^m - \theta^\circ\|_{l_2}^2 > 4 \left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n} \right] \right\} \right] \leq \exp\left[-\frac{m}{9L} \right].$$

In addition to that, thanks to the definitions of G_n^- and G_n^+, the monotonicity of $\mathfrak{b}_m$ and $\frac{m\overline{\Lambda}_m}{n}$, on the one hand we have, for all m in $[\![G_n^-, m_n^\circ]\!]$:

$$\begin{aligned} \frac{m\overline{\Lambda}_m}{n} &\leq \frac{m_n^\circ \overline{\Lambda}_{m_n^\circ}}{n} \leq \Phi_n^\circ, \\ \mathfrak{b}_m &\leq 9L\Phi_n^\circ; \end{aligned}$$

and on the other hand, thanks to ASSUMPTION 1.2, with $D^\circ := \left\lceil \frac{3}{\kappa^\circ} + 1 \right\rceil$, then we have for all m in $[\![m_n^\circ, G_n^+]\!]$:

$$\begin{aligned} m &\leq \frac{3\Phi_n^\circ n}{\Lambda_{m_n^\circ}} + m_n^\circ \leq \frac{3nm_n^\circ \overline{\Lambda}_{m_n^\circ}}{\Lambda_{m_n^\circ} n\kappa^\circ} + m_n^\circ \leq \left(\frac{3}{\kappa^\circ} + 1 \right) m_n^\circ \leq D^\circ m_n^\circ, \\ \mathfrak{b}_m &\leq \mathfrak{b}_{m_n^\circ} \leq \Phi_n^\circ. \end{aligned}$$

Using $m \leq D^\circ m_n^\circ$ and ASSUMPTION 1.1 we have

$$\overline{\Lambda}_m \leq \Lambda_m \leq \Lambda_{D^\circ m_n^\circ} \leq \Lambda_{D^\circ} \Lambda_{m_n^\circ} \leq \Lambda_{D^\circ} L \overline{\Lambda}_{m_n^\circ}.$$

So finally we have for all m in $[\![G_n^-, G_n^+]\!]$, with $K^\circ \geq 4L \cdot [9 \vee D^\circ \Lambda_{D^\circ}]$:

$$\begin{aligned} \mathfrak{b}_m &\leq 9L\Phi_n^\circ, \\ \frac{m\overline{\Lambda}_m}{n} &\leq D^\circ \Lambda_{D^\circ} L \cdot \frac{m_n^\circ \overline{\Lambda}_{m_n^\circ}}{n} \leq D^\circ \Lambda_{D^\circ} L \cdot \Phi_n^\circ, \\ \left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n} \right] &\leq L \left[9 \vee D^\circ \Lambda_{D^\circ} \right] \Phi_n^\circ, \\ 4 \left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n} \right] &\leq K^\circ \Phi_n^\circ. \end{aligned}$$

Which leads us to the upper bound LES DEUX FORMULES DE SOMME D'exp NE SEMBLENT PAS EN ACCORD ?

$$\sum_{m=G_n^-}^{G_n^+} B_{3,m} \leq \sum_{m=G_n^-}^{G_n^+} \exp\left[-\frac{m}{9L} \right] \leq 18L \exp\left[-\frac{G_n^-}{9L} \right].$$

Finally, we can conclude :

$$\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\left(K^\circ\right)^{-1}\Phi^\circ_n \leq \|\boldsymbol{\theta}^M - \theta^\circ\|^2 \leq K^\circ \Phi^\circ_n\right)\right]$$
$$\geq 1 - 2\exp\left[-\frac{5m^\circ_n}{9L} + \log\left(G_n\right)\right] - 2\exp\left[-\frac{7m^\circ_n}{9} + \log\left(G_n\right)\right] - 2K^\circ\exp\left[-\frac{G^-_n}{K^\circ}\right] - 18L\exp\left[-\frac{G^-_n}{9L}\right].$$

E Proof for THEOREM 4.3

DEFINITION E.1 Define the following quantities :

$$G^{\star-}_n := \min\left\{m \in [\![1, m^\star_n]\!] : \quad \mathfrak{b}_m \leq 9\left(1 \vee r\right) L\Psi^\star_n\right\},$$
$$G^{\star+}_n := \max\left\{m \in [\![m^\star_n, G_n]\!] : \frac{m - m^\star_n}{n} \leq 3\Lambda^{-1}_{m^\star_n}\left(1 \vee r\right)\Psi^\star_n\right\}.$$

PROPOSITION E.1 Under ASSUMPTION 1.1, we have the following concentration inequalities for the threshold hyper parameter :

$$\mathbb{P}^n_{\theta^\circ}\left[\widehat{m} > G^{\star+}_n\right] \leq \exp\left[-\frac{5\left(1 \vee r\right)m^\star_n}{9L} + \log\left(G_n\right)\right],$$
$$\mathbb{P}^n_{\theta^\circ}\left[\widehat{m} < G^{\star-}_n\right] \leq \exp\left[-\frac{7\left(1 \vee r\right)m^\star_n}{9} + \log\left(G_n\right)\right].$$

E.1 Proof of PROPOSITION E.1

First, let's proof the first inequality. Use the fact that :

$$\begin{aligned}
&\mathbb{P}^n_{\theta^\circ}\left[G^{\star+}_n < \widehat{m} \leq G_n\right]\\
&= \mathbb{P}^n_{\theta^\circ}\left[\forall l \in [\![1, G^{\star+}_n]\!], \quad \frac{3\widehat{m}}{n} - \sum_{j=1}^{\widehat{m}} Y_j^2 < \frac{3l}{n} - \sum_{j=1}^{l} Y_j^2\right]\\
&\leq \mathbb{P}^n_{\theta^\circ}\left[\exists m \in [\![G^{\star+}_n + 1, G_n]\!] : \quad \frac{3m}{n} - \sum_{j=1}^{m} Y_j^2 < \frac{3m^\star_n}{n} - \sum_{j=1}^{m^\star_n} Y_j^2\right]\\
&\leq \sum_{m=G^{\star+}_n+1}^{G_n} \mathbb{P}^n_{\theta^\circ}\left[\frac{3m}{n} - \sum_{j=1}^{m} Y_j^2 < \frac{3m^\star_n}{n} - \sum_{j=1}^{m^\star_n} Y_j^2\right]\\
&\leq \sum_{m=G^{\star+}_n+1}^{G_n} \mathbb{P}^n_{\theta^\circ}\left[0 < 3\frac{m^\star_n - m}{n} + \sum_{j=m^\star_n+1}^{m} Y_j^2\right]
\end{aligned}$$

We will now use LEMMA C.1. For this purpose, define then for all m in $[\![G^{\star+}_n + 1, G_n]\!]$: $S_m := \sum_{j=m^\star_n+1}^{m} Y_j^2$, we then have $\mu_m := \mathbb{E}^n_{\theta^\circ}\left[S_m\right] = \frac{m-m^\star_n}{n} + \sum_{j=m^\star_n+1}^{m}\left(\theta^\circ_j\lambda_j\right)^2$, $\alpha_j^2 := \mathbb{E}^n_{\theta^\circ}\left[Y_j\right]^2 = \left(\theta^\circ_j\lambda_j\right)^2$ and $\beta_j^2 := \mathbb{V}^n_{\theta^\circ}\left[Y_j\right] = \frac{1}{n}$.

Now we note, using the definition of $\Theta^{\mathfrak{a}}(r)$

$$\begin{aligned}
\sum_{j=m_n^\star+1}^{m} \alpha_j^2 &= \sum_{j=m_n^\star+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2 \\
&\leq \Lambda_{m_n^\star}^{-1} \sum_{j=m_n^\star+1}^{m} \left(\theta_j^\circ\right)^2 \\
&\leq \Lambda_{m_n^\star}^{-1} \mathfrak{b}_{m_n^\star} \\
&\leq \Lambda_{m_n^\star}^{-1} \left(1 \vee r\right) \Psi_n^\star &&=: r_m \\
\sum_{j=m_n^\star}^{m} \beta_j^2 &= \frac{m - m_n^\star}{n} &&=: v_m \\
\max_{j \in [\![m_n^\star, m]\!]} \beta_j &= \frac{1}{n} &&=: t_m.
\end{aligned}$$

Hence, we have, for all m in $[\![G_n^{\star+} + 1, G_n]\!]$

$$\begin{aligned}
&\mathbb{P}_{\theta^\circ}^n \left[\sum_{j=m_n^\star+1}^{m} Y_j^2 - 3\frac{m - m_n^\star}{n} > 0 \right] \\
&= \mathbb{P}_{\theta^\circ}^n \left[S_m - \frac{m - m_n^\star}{n} > 2\frac{m - m_n^\star}{n} \right] \\
&= \mathbb{P}_{\theta^\circ}^n \left[S_m - \frac{m - m_n^\star}{n} - \sum_{j=m_n^\star+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2 > 2\frac{m - m_n^\star}{n} - \sum_{j=m_n^\star+1}^{m} \left(\theta_j^\circ \lambda_j\right)^2 \right] \\
&\leq \mathbb{P}_{\theta^\circ}^n \left[S_m - \mu_m > 2\frac{m - m_n^\star}{n} - \Lambda_{m_n^\star}^{-1} \mathfrak{b}_{m_n^\star} \right] \\
&\leq \mathbb{P}_{\theta^\circ}^n \left[S_m - \mu_m > 2\frac{m - m_n^\star}{n} - \Lambda_{m_n^\star}^{-1} \left(1 \vee r\right) \Psi_n^\star \right].
\end{aligned}$$

Using the definition of $G_n^{\star+}$, we have $\frac{m - m_n^\star}{n} > 3\Lambda_{m_n^\star}^{-1} \left(1 \vee r\right) \Psi_n^\star$.
Hence, we can write :

$$
\begin{aligned}
\mathbb{P}^n_{\theta^\circ}&\left[\sum_{j=m_n^\star}^{m} Y_j^2 - 3\frac{m-m_n^\star}{n} > 0\right] \\
&\leq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m > \frac{m-m_n^\star}{n} + 2\Lambda_{m_n^\star}^{-1}\left(1\vee r\right)\Psi_n^\star\right] \\
&\leq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m > v_m + 2r_m\right] \\
&\leq \exp\left[-n\frac{\frac{m-m_n^\star}{n} + 2\Lambda_{m_n^\star}^{-1}\left(1\vee r\right)\Psi_n^\star}{9}\right] \\
&\leq \exp\left[-n\frac{5\Lambda_{m_n^\star}^{-1}\left(1\vee r\right)\Psi_n^\star}{9}\right] \\
&\leq \exp\left[-\frac{5\left(1\vee r\right)m_n^\star}{9L}\right].
\end{aligned}
$$

Finally we can conclude that

$$
\mathbb{P}^n_{\theta^\circ}\left[G_n^{\star+} < \widehat{m} \leq G_n\right] \leq \exp\left[-\frac{5\left(1\vee r\right)m_n^\star}{9L} + \log\left(G_n\right)\right].
$$

We now prove the second inequality.
We begin by writing the same kind of inclusion of events as for the first inequality :

$$
\begin{aligned}
\mathbb{P}^n_{\theta^\circ}\left[1 \leq \widehat{m} < G_n^{\star-}\right] =&\mathbb{P}^n_{\theta^\circ}\left[\forall m \in [\![G_n^-, G_n]\!], \quad \Upsilon\left(\widehat{m}\right) + \mathrm{pen}\left(\widehat{m}\right) < \Upsilon\left(m\right) + \mathrm{pen}\left(m\right)\right] \\
\leq&\mathbb{P}^n_{\theta^\circ}\left[\exists m \in [\![1, G_n^{\star-} - 1]\!], \quad \Upsilon\left(m\right) + \mathrm{pen}\left(m\right) < \Upsilon\left(m_n^\star\right) + \mathrm{pen}\left(m_n^\star\right)\right] \\
\leq& \sum_{m=1}^{G_n^{\star-}} \mathbb{P}^n_{\theta^\circ}\left[\Upsilon\left(m\right) + \mathrm{pen}\left(m\right) < \Upsilon\left(m_n^\star\right) + \mathrm{pen}\left(m_n^\star\right)\right] \\
\leq& \sum_{m=1}^{G_n^{\star-}} \mathbb{P}^n_{\theta^\circ}\left[\sum_{j=m+1}^{m_n^\star} Y_j^2 < 3\frac{m_n^\star - m}{n}\right].
\end{aligned}
$$

The LEMMA 3.1 steps in again, define $S_m := \sum_{j=m+1}^{m_n^\star} Y_j^2$ and we want to control the concentration of this sum, hence we take the following notations :

$$
\begin{aligned}
\mu_m :=& \mathbb{E}^n_{\theta^\circ}\left[S_m\right] \\
=& \frac{m^\star_n - m}{n} + \sum_{j=m+1}^{m^\star_n} \left(\theta^\circ_j \lambda_j\right)^2 \\
r_m :=& \sum_{j=m+1}^{m^\star_n} \left(\theta^\circ_j \lambda_j\right)^2 \\
v_m :=& \frac{m^\star_n - m}{n} \\
t_m :=& \frac{1}{n}.
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
&\mathbb{P}^n_{\theta^\circ}\left[S_m < 3\frac{m^\star_n - m}{n}\right] \\
&= \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < 3\frac{m^\star_n - m}{n} - \frac{m^\star_n - m}{n} - \sum_{j=m+1}^{m^\star_n} \left(\theta^\circ_j \lambda_j\right)^2\right] \\
&= \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < 3\frac{m^\star_n - m}{n} - \frac{2}{3}\frac{m^\star_n - m}{n} - \frac{1}{3}\sum_{j=m+1}^{m^\star_n} \left(\theta^\circ_j \lambda_j\right)^2 - \frac{1}{3}\left[v_m + 2r_m\right]\right] \\
&\leq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < \frac{7}{3}\frac{m^\star_n - m}{n} - \frac{1}{3}\Lambda^{-1}_{m^\star_n}\sum_{j=m+1}^{m^\star_n} \left(\theta^\circ_j\right)^2 - \frac{1}{3}\left[v_m + 2r_m\right]\right] \\
&\leq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < -\frac{1}{3}\left[v_m + 2r_m\right] + \frac{7}{3}\frac{m^\star_n}{n} + \frac{1}{3}\Lambda^{-1}_{m^\star_n}\mathfrak{b}_{m^\star_n} - \frac{1}{3}\Lambda^{-1}_{m^\star_n}\mathfrak{b}_m\right] \\
&\leq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < -\frac{1}{3}\left[v_m + 2r_m\right] + 3\left(1 \vee r\right)\Psi^\star_n\overline{\Lambda}^{-1}_{m^\star_n} - \frac{1}{3}\Lambda^{-1}_{m^\star_n}\mathfrak{b}_m\right]
\end{aligned}
$$

now, using the definition of G^-_n, we have $\mathfrak{b}_m > 9L\left(1 \vee r\right)\Psi^\star_n$ so

$$
\begin{aligned}
&\mathbb{P}^n_{\theta^\circ}\left[S_m < 3\frac{m_n^\star - m}{n}\right] \\
&\leq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m < -\frac{1}{3}\left[v_m + 2r_m\right]\right] \\
&\leq \exp\left[-\frac{n}{36}\left(\frac{m_n^\star - m}{n} + 2\sum_{j=m+1}^{m_n^\star}\left(\theta_j^\circ\lambda_j\right)^2\right)\right] \\
&\leq \exp\left[-n\frac{\frac{m_n^\star - m}{n} + 2\Lambda_{m_n^\star}^{-1}\mathfrak{b}_m - 2\Lambda_{m_n^\star}^{-1}\mathfrak{b}_{m_n^\star}}{36}\right] \\
&\leq \exp\left[-n\frac{16L\left(1\vee r\right)\Psi_n^\star\Lambda_{m_n^\star}^{-1}}{36}\right] \\
&\leq \exp\left[-\frac{4\left(1\vee r\right)m_n^\star}{9}\right]
\end{aligned}
$$

Which in turn implies

$$
\mathbb{P}^n_{\theta^\circ}\left[1 \leq \widehat{m} < G_n^{\star-}\right] \leq \exp\left[-\frac{4\left(1\vee r\right)m_n^\star}{9} + \log\left(G_n\right)\right].
$$

E.2 Proof of the final statement

The L^2 risk can be written :

$$
\mathbb{E}^n_{\theta^\circ}\left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] = \mathbb{E}^n_{\theta^\circ}\left[\sum_{j=1}^{G_n}\left(\overline{\theta}_j - \theta_j^\circ\right)^2\right] + \mathbb{E}^n_{\theta^\circ}\left[\sum_{j=G_n+1}^{\infty}\left(\theta_j^\circ\right)^2\right].
$$

Together with

$$
\forall j \in [\![1, G_n]\!], \quad \overline{\theta}_j^{\widehat{m}} - \theta_j^\circ = \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ\right)\mathbb{1}_{\{\widehat{m}\in[\![j, G_n]\!]\}} + \theta_j^\circ\mathbb{1}_{\{\widehat{m}\in[\![1, j-1]\!]\}},
$$

implies that

$$
\begin{aligned}
\mathbb{E}^n_{\theta^\circ}\left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] \leq & \underbrace{\sum_{j=1}^{G_n}\mathbb{E}^n_{\theta^\circ}\left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ\right)^2\mathbb{1}_{\{\widehat{m}\in[\![j, G_n]\!]\}}\right]}_{=:A} \\
& + \underbrace{\sum_{j=1}^{G_n}\mathbb{E}^n_{\theta^\circ}\left[\left(\theta_j^\circ\right)^2\mathbb{1}_{\{\widehat{m}\in[\![1, j-1]\!]\}}\right]}_{=:B} + \underbrace{\sum_{j=G_n+1}^{\infty}\left(\theta_j^\circ\right)^2}_{=:C}.
\end{aligned}
$$

We will now control each of the three parts of the sum.

First, consider A and let be some positive real number p to be specified later.

Then we can write

$$
\begin{aligned}
A \leq & \sum_{j=1}^{G_n^{\star+}} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \sum_{j=G_n^{\star+}+1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \mathbb{1}_{\{\widehat{m} \in [\![j, G_n]\!]\}} \right] \\
\leq & \sum_{j=1}^{G_n^{\star+}} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \sum_{j=G_n^{\star+}+1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]\}} \right] \\
\leq & \sum_{j=1}^{G_n^{\star+}} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \sum_{j=G_n^{\star+}+1}^{G_n} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]\}} \right] \\
& - p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]} \right] + p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]} \right] \\
\leq & \sum_{j=1}^{G_n^{\star+}} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right] + \mathbb{E}_{\theta^\circ} \left[\left(\sum_{j=G_n^{\star+}+1}^{G_n} \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 - p \right) \mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]\}} \right] \\
& + p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]} \right] \\
\leq & \underbrace{\sum_{j=1}^{G_n^{\star+}} \mathbb{E}_{\theta^\circ}^n \left[\left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 \right]}_{=:A_3} + \underbrace{\mathbb{E}_{\theta^\circ}^n \left[\left(\sum_{j=G_n^{\star+}+1}^{G_n} \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2 - p \right)_+ \right]}_{=:A_1} + \underbrace{p \mathbb{E}_{\theta^\circ}^n \left[\mathbb{1}_{\{\widehat{m} \in [\![G_n^{\star+}, G_n]\!]} \right]}_{=:A_2}.
\end{aligned}
$$

First, we will control A_1 using LEMMA C.2.
The goal is to give p a value that is large enough to control this object but small enough so $p \cdot \mathbb{P}_{\theta^\circ}^n \left[G_n^{\star+} < \widehat{m} \leq G_n \right]$ is still for the most $\Psi_n^\star$.
Define $S_n := \sum_{j=G_n^{\star+}+1}^{G_n} \left(\frac{Y_j}{\lambda_j} - \theta_j^\circ \right)^2$.
We have, for all j in $[\![G_n^{\star+} + 1, G_n]\!]$,

$$
\frac{Y_j}{\lambda_j} - \theta_j^\circ \sim \mathcal{N} \left(0, \frac{\Lambda_j}{n} \right),
$$

so $\mathbb{E}_{\theta^\circ}^n \left[S_n \right] = \frac{1}{n} \sum_{j=G_n^{\star+}+1}^{G_n} \Lambda_j$.
And define

$$
\begin{aligned}
t_m :=& \Lambda_1 && \geq \frac{\Lambda_{G_n}}{n} && \geq \max_{j \in [\![G_n^{\star+}+1, G_n]\!]} \frac{\Lambda_j}{n}, \\
v_m :=& G_n \Lambda_1 && \geq \frac{G_n \Lambda_{G_n}}{n} && \geq \frac{1}{n} \sum_{j=G_n^{\star+}+1}^{G_n} \Lambda_j.
\end{aligned}
$$

We can take $p = \mathbb{E}^n_{\theta^\circ}[S_n] + 3v_m$ which gives

$$\begin{aligned}
A_1 =& \mathbb{E}^n_{\theta^\circ}\left[(S_n - \mathbb{E}^n_{\theta^\circ}[S_n] - 3v_m)_+\right]\\
\leq& 6\Lambda_1 \exp\left[-\frac{G_n}{2}\right]\\
\leq& 6\Lambda_1 \exp\left[-\frac{nG_n}{2n}\right]\\
\leq& 6\Lambda_1 \exp\left[-n\frac{3\Lambda^{-1}_{m^\star_n}(1\vee r)\Psi^\star_n}{2} - \frac{m^\star_n}{2}\right]\\
A_1 \leq& 6\Lambda_1 \exp\left[-\frac{2m^\star_n}{L}\right].
\end{aligned}$$

Thanks to PROPOSITION E.1 it is easily shown that

$$A_2 \leq 4\Lambda_1 \exp\left[-\frac{5m^\star_n}{9L} + 2\log(G_n)\right].$$

Finally, we control A_3. Using the definition of $G^{\star+}_n$ we have

$$\begin{aligned}
A_3 =& \sum_{j=1}^{G^{\star+}_n} \mathbb{E}^n_{\theta^\circ}\left[\left(\overline{\theta}_j - \theta^\circ_j\right)^2\right]\\
=& \sum_{j=1}^{G^{\star+}_n} \frac{\Lambda_j}{n}\\
=& \frac{1}{n} G^{\star+}_n \overline{\Lambda}_{G^{\star+}_n}\\
A_3 \leq& \frac{\overline{\Lambda}_{G^{\star+}_n}}{\Lambda_{m^\star_n}} \Psi^\star_n \left(3(1\vee r) + \frac{\Lambda_{m^\star_n}}{\overline{\Lambda}_{m^\star_n}}\right).
\end{aligned}$$

Note that, using ASSUMPTION 1.1 and the definition of $G^{\star+}_n$, we have that $\frac{\overline{\Lambda}_{G^{\star+}_n}}{\Lambda_{m^\star_n}}$ is bounded for n large enough; indeed with $D^\star := \left\lceil \frac{3(1\vee r)}{\kappa^\star} + 1 \right\rceil$,

$$\begin{aligned}
&G^{\star+}_n \leq \frac{3(1\vee r)\Psi^\star_n n}{\Lambda_{m^\star_n}} + m^\star_n \leq \frac{3(1\vee r) m^\star_n \overline{\Lambda}_{m^\star_n} n}{n\Lambda_{m^\star_n}\kappa^\star} + m^\star_n \leq \left(\frac{3(1\vee r)}{\kappa^\star} + 1\right) m^\star_n \leq D^\star m^\star_n\\
&\Rightarrow \frac{\overline{\Lambda}_{G^{\star+}_n}}{\Lambda_{m^\star_n}} \leq \Lambda_{D^\star}.
\end{aligned}$$

Hence, we have

$$A \leq 6\Lambda_1 \exp\left[-\frac{2m^\star_n}{L}\right] + 4\Lambda_1 \exp\left[-\frac{5m^\star_n}{9L} + 2\log(G_n)\right] + 4\Lambda_{D^\star} L(1\vee r)\Psi^\star_n.$$

Now we control B. We use a similar decomposition to the one used for A :

$$
\begin{aligned}
B &\leq \sum_{j=G_n^{\star-}+1}^{G_n} \left(\theta_j^\circ\right)^2 + \sum_{j=1}^{G_n^{\star-}} \mathbb{E}_{\theta^\circ}^n \left[\left(\theta_j^\circ\right)^2 \mathbb{1}_{\{\widehat{m}\in[\![1,j-1]\!]\}}\right] \\
&\leq \sum_{j=G_n^{\star-}+1}^{G_n} \left(\theta_j^\circ\right)^2 + \sum_{j=1}^{G_n^{\star-}} \mathbb{E}_{\theta^\circ}^n \left[\left(\theta_j^\circ\right)^2 \mathbb{1}_{\left\{\widehat{m}\in[\![1,G_n^{\star-}-1]\!]\right\}}\right] \\
B &\leq \underbrace{\sum_{j=G_n^{\star-}+1}^{G_n} \left(\theta_j^\circ\right)^2}_{=:B_1} + \underbrace{\sum_{j=1}^{G_n^{\star-}} \left(\theta_j^\circ\right)^2 \mathbb{P}_{\theta^\circ}^n \left[1 \leq \widehat{m} < G_n^{\star-}\right]}_{=:B_2}.
\end{aligned}
$$

First, notice that $B_1 + C = \mathfrak{b}_{G_n^{\star-}} \leq 9L\left(1 \vee r\right)\Psi_n^\star$ by the definition of $G_n^{\star-}$.

To control B_2, we use the definition of the Sobolev ellipsoid and the PROPOSITION E.1:

$$
\begin{aligned}
B_2 =& \mathbb{P}_{\theta^\circ}^n \left[1 \leq \widehat{m} < G_n^{\star-}\right] \sum_{j=1}^{G_n^{\star-}} \left(\theta_j^\circ\right)^2 \\
\leq& \exp\left[-\frac{7\left(1 \vee r\right) m_n^\star}{9} + \log\left(G_n\right)\right] \cdot \sum_{j=1}^{G_n^{\star-}} \frac{\mathfrak{a}_j}{\mathfrak{a}_j} \left(\theta_j^\circ\right)^2 \\
\leq& \exp\left[-\frac{7\left(1 \vee L^\circ\right) m_n^\star}{9} + \log\left(G_n\right)\right] \cdot \mathfrak{a}_1 \sum_{j=1}^{G_n^{\star-}} \frac{1}{\mathfrak{a}_j} \left(\theta_j^\circ\right)^2 \\
\leq& \exp\left[-\frac{7m_n^\star}{9} + \log\left(G_n\right)\right] \cdot \mathfrak{a}_1 r
\end{aligned}
$$

So we have

$$
B + C \leq 9L\left(1 \vee r\right)\Psi_n^\star + \mathfrak{a}_1 L^\circ \cdot \exp\left[-\frac{7m_n^\star}{9} + \log\left(G_n\right)\right].
$$

Which leads to :

$$
\begin{aligned}
\mathbb{E}_{\theta^\circ}^n \left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] \leq& L\left(1 \vee L^\circ\right)\left(4\Lambda_{D^\star} + 9\right)\Psi_n^\star + \\
&\left(6\Lambda_1 \exp\left[-\frac{2m_n^\star}{L} - \log\left(\Psi_n^\star\right)\right] + 4\Lambda_1 \exp\left[-\frac{5m_n^\star}{9L} + \log\left(\frac{G_n^2}{\Psi_n^\star}\right)\right] + \right. \\
&\left. \mathfrak{a}_1 r \cdot \exp\left[-\frac{7m_n^\star}{9} + \log\left(\frac{G_n}{\Psi_n^\star}\right)\right]\right)\Psi_n^\star,
\end{aligned}
$$

which proves that there exist K such that, for n large enough,

$$
\mathbb{E}_{\theta^\circ}^n \left[\left\|\overline{\theta}^{\widehat{m}} - \theta^\circ\right\|^2\right] \leq C\Psi_n^\star.
$$

Add remark stating that the bound is uniform over the ellipsoid hence rate for the maximal risk?

F Proof of THEOREM 4.4

PROPOSITION F.1 Under ASSUMPTION 1.1, we have, for all m in $[\![1, G_n]\!]$ and c greater than $\frac{3}{2}$,

$$\mathbb{P}^n_{\theta^\circ}\left[\left\|\overline{\theta}^m - \theta^\circ\right\|^2 > 4c\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right]\right] \leq \exp\left[-\frac{cm}{6L}\right].$$

F.1 Proof of PROPOSITION F.1

Let be m in $[\![1, G_n]\!]$ and note that

$$\|\overline{\theta}^m - \theta^\circ\|^2 = \sum_{j=1}^{m}\left(\frac{Y_j}{\lambda_j} - \theta^\circ_j\right)^2 + \mathfrak{b}_m,$$

hence, we will use LEMMA C.1.
We then define

$$\begin{aligned}
&S_m := \sum_{j=1}^{m}\left(\frac{Y_j}{\lambda_j} - \theta^\circ_j\right)^2,\\
&\mu_m := \mathbb{E}^n_{\theta^\circ}\left[S_m\right] = \frac{m\overline{\Lambda}_m}{n},\\
\forall j \in [\![1, m]\!] \quad &\beta_j^2 := \mathbb{V}^n_{\theta^\circ}\left[\frac{Y_j}{\lambda_j} - \theta^\circ_j\right] = \frac{\Lambda_j}{n},\\
\forall j \in [\![1, m]\!] \quad &\alpha_j^2 := \mathbb{E}^n_{\theta^\circ}\left[\frac{Y_j}{\lambda_j} - \theta^\circ_j\right] = 0,\\
&v_m := \sum_{j=1}^{m}\beta_j^2 = \frac{m\overline{\Lambda}_m}{n},\\
&t_m := \max_{j\in[\![1,m]\!]}\beta_j^2 = \frac{\Lambda_m}{n}.
\end{aligned}$$

We then control the concentration of S_m, define c a constant greater than $\frac{3}{2}$:

$$\begin{aligned}
\exp\left[-\frac{cm}{6L}\right] &\geq \exp\left[-\frac{cnm\overline{\Lambda}_m}{6n\Lambda_m}\right]\\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m - \mu_m \geq c\frac{m\overline{\Lambda}_m}{n}\right]\\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m \geq 2c\frac{m\overline{\Lambda}_m}{n}\right]\\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \geq 2c\frac{m\overline{\Lambda}_m}{n} + \mathfrak{b}_m\right]\\
&\geq \mathbb{P}^n_{\theta^\circ}\left[S_m + \mathfrak{b}_m \geq 4c\left(\frac{m\overline{\Lambda}_m}{n} \vee \mathfrak{b}_m\right)\right].
\end{aligned}$$

F.2 Proof of the final statement

By the the total probability formula, we have :

$$\begin{aligned}
&\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\|\boldsymbol{\theta}^M-\theta^\circ\|^2 \leq K^\star\Psi^\star_n\right)\right]\\
&= 1-\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(K^\star\Psi^\star_n < \|\boldsymbol{\theta}^M-\theta^\circ\|^2\right)\right].
\end{aligned}$$

Hence, we will control $\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(K^\star\Psi^\star_n < \|\boldsymbol{\theta}^M-\theta^\circ\|^2\right)\right]$.
We can write :

$$\begin{aligned}
\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(K^\star\Psi^\star_n < \|\boldsymbol{\theta}-\theta^\circ\|^2\right)\right] &= \sum_{m=1}^{G_n}\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\left\{K^\star\Psi^\star_n < \|\boldsymbol{\theta}^M-\theta^\circ\|^2\right\}\cap\{M=m\}\right)\right]\\
&\leq \sum_{m=1}^{G_n^{\star-}-1}\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{M|Y^n}\left(\{M=m\}\right)\right] + \sum_{m=G_n^{\star+}+1}^{G_n}\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\{M=m\}\right)\right]\\
&+ \sum_{m=G_n^{\star-}}^{G_n^{\star+}}\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y,M=m}\left(\left\{K^\star\Psi^\star_n < \|\boldsymbol{\theta}^M-\theta^\circ\|^2\right\}\right)\right]\\
&\leq \underbrace{\sum_{m=1}^{G_n^{\star-}-1}\mathbb{P}^n_{\theta^\circ}\left[\{\widehat{m}=m\}\right]}_{=:A} + \underbrace{\sum_{m=G_n^{\star+}+1}^{G_n}\mathbb{P}^n_{\theta^\circ}\left[\{\widehat{m}=m\}\right]}_{=:B} + \sum_{m=G_n^{\star-}}^{G_n^{\star+}}\underbrace{\mathbb{P}^n_{\theta^\circ}\left[\left\{K^\star\Psi^\star_n < \left\|\overline{\theta}^m-\theta^\circ\right\|^2\right\}\right]}_{=:C_m}.
\end{aligned}$$

As previously, A and B are controlled in PROPOSITION E.1. Hence, we now control $\sum_{m=G_n^{\star-}}^{G_n^{\star+}} C_m$.

Start by noting that for all m in $[\![G_n^{\star-}, m_n^\star]\!]$, we have $\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right] \leq 9L\,(1\vee r)\,\Psi^\star_n$ and, if we define $D^\star := \left\lceil\frac{3(1\vee r)}{\kappa^\star}+1\right\rceil$ for all m in $[\![m_n^\star, G_n^{\star+}]\!]$, we have $\left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n}\right] \leq D^\star\Lambda_{D^\star}\,(1\vee r)\,\Psi^\star_n$.
Using PROPOSITION F.1, we have that, for all m in $[\![G_n^{\star-}, G_n^{\star+}]\!]$ and c greater than $\frac{3}{2}$:

$$\mathbb{P}^n_{\theta^\circ}\left[\left\{\|\overline{\theta}^m-\theta^\circ\|^2 > 4c\,(9L\vee D^\star\Lambda_{D^\star})\,(1\vee r)\,\Psi^\star_n\right\}\right] \leq \exp\left[-\frac{cm}{6L}\right].$$

Hence, we set $K^\star := 6\,(9L\vee D^\star\Lambda_{D^\star})\,(1\vee r)$, which leads us to the upper bound LES DEUX FORMULES DE SOMME D'exp NE SEMBLENT PAS EN ACCORD ?

$$\sum_{m=G_n^{\star-}}^{G_n^{\star+}} C_m \leq \sum_{m=G_n^{\star-}}^{G_n^{\star+}}\exp\left[-\frac{m}{4L}\right] \leq 8L\exp\left[-\frac{G_n^{\star-}}{4L}\right].$$

Finally, we can conclude :

$$
\begin{aligned}
&\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\|\boldsymbol{\theta}^M-\theta^\circ\|^2\leq K^\star\Psi^\star_n\right)\right]\\
&\quad\geq 1-2\exp\left[-\frac{5m^\star_n}{9L}+\log\left(G_n\right)\right]-2\exp\left[-\frac{7m^\star_n}{9}+\log\left(G_n\right)\right]-8L\exp\left[-\frac{G^{\star -}_n}{4L}\right].
\end{aligned}
$$

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 it is also, by definition of $G^{\star -}_n$, we have $\mathfrak{b}_{G^{\star -}_n} \leq 9\left(1\vee r\right)L\Psi^\star_n$ which leads to the conclusion that for all m greater than $G^{\star -}_{n^\circ}$, $\mathfrak{b}_m = 0$. Hence, for all m greater than $G^{\star -}_{n^\circ}$, we can write

$$
\frac{K^\star\cdot\Psi^\star_n}{\left[\mathfrak{b}_m\vee\frac{m\overline{\Lambda}_m}{n}\right]}=\frac{K^\star\Psi^\star_n}{\left[\frac{m\overline{\Lambda}_m}{n}\right]}\geq\frac{K^\star\frac{m^\star_n\overline{\Lambda}_{m^\star_n}}{n}}{\left[\frac{m\overline{\Lambda}_m}{n}\right]}\geq\frac{K^\star m^\star_n}{Lm}\geq 9D^\star\Lambda_{D^\star}\left(1\vee r\right)\frac{m^\star_n}{m}\geq 1.
$$

Hence, we set $c := 18D^\star\Lambda_{D^\star}\left(1\vee r\right)\frac{m^\star_n}{m}\geq\frac{3}{2}$ and can write in this case for all n smaller than n° :

$$
\begin{aligned}
\sum_{m=G^{\star -}_n}^{G^{\star +}_n}\mathbb{P}^n_{\theta^\circ}\left[K^\star\Psi^\star_n<\left\|\overline{\theta}^m-\theta^\circ\right\|^2\right]&\leq\sum_{m=G^{\star -}_n}^{G^{\star +}_n}\mathbb{P}^n_{\theta^\circ}\left[4c\left[\mathfrak{b}_m\vee\frac{m\overline{\Lambda}_m}{n}\right]<\left\|\overline{\theta}^m-\theta^\circ\right\|^2\right]\\
&\leq\sum_{m=G^{\star -}_n}^{G^{\star +}_n}\exp\left[-\frac{cm}{6L}\right]\\
&\leq\sum_{m=G^{\star -}_n}^{G^{\star +}_n}\exp\left[-\frac{3D^\star\Lambda_{D^\star}\left(1\vee r\right)m^\star_n}{L}\right]\\
&\leq\exp\left[-\frac{3D^\star\Lambda_{D^\star}\left(1\vee r\right)m^\star_n}{L}+\log\left(G_n\right)\right].
\end{aligned}
$$

We can hence conclude that

$$
\begin{aligned}
\mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^{n,(\infty)}_{\boldsymbol{\theta}^M|Y^n}\left(\|\boldsymbol{\theta}^M-\theta^\circ\|^2\leq K^\star\Psi^\star_n\right)\right]&>1-\exp\left[-\frac{3D^\star\Lambda_{D^\star}\left(1\vee r\right)m^\star_n}{L}+\log\left(G_n\right)\right]\\
&-2\exp\left[-\frac{5m^\star_n}{9L}+\log\left(G_n\right)\right]-2\exp\left[-\frac{7m^\star_n}{9}+\log\left(G_n\right)\right].
\end{aligned}
$$

Hence, we have shown here that $\Psi^\star_n$ is an upper bound for the contraction rate under the quadratic risk. We will now use this to proof that it is also for the minimax risk.

Note that $K^\star\,\Psi^\star_n\geq 4\left[\mathfrak{b}_m\vee\frac{m\overline{\Lambda}_m}{n}\right]$ 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

$$
K_n\Psi^\star_n\geq 4\frac{K_n}{K^\star}\left[\mathfrak{b}_m\vee\frac{m\overline{\Lambda}_m}{n}\right].
$$

So, if we define $\tilde{n}^\circ$, the smallest integer such that $\frac{K_n}{K^\star} \geq 1$, we can apply PROPOSITION F.1 and we have

$$\sum_{m=G_n^{\star-}}^{G_n^{\star+}} \mathbb{P}_{\theta^\circ}^n \left[K_n \Psi_n^\star < \left\| \overline{\theta}^m - \theta^\circ \right\|^2 \right] \leq \sum_{m=G_n^{\star-}}^{G_n^{\star+}} \mathbb{P}_{\theta^\circ}^n \left[4\frac{K_n}{K^\star} \left[\mathfrak{b}_m \vee \frac{m\overline{\Lambda}_m}{n} \right] < \left\| \overline{\theta}^m - \theta^\circ \right\|^2 \right]$$
$$\leq \sum_{m=G_n^{\star-}}^{G_n^{\star+}} \exp\left[-\frac{4K_n m}{9K^\star L} \right]$$
$$\leq \exp\left[-\frac{4K_n}{9K^\star L} \right].$$

VERIFIER FORMULE exp

We hence here have a uniform upper bound for the maximal risk which concludes the proof.

References

BUNKE, O. and JOHANNES, J. (2005). Selfinformative limits of bayes estimates and generalized maximum likelihood. *Statistics*, 39(6):483–502.

JOHANNES, J., SIMONI, A., and SCHENK, R. (2016). Adaptive bayesian estimation in indirect gaussian sequence space model.