FRONTIERS TO THE LEARNING OF NONPARAMETRIC HIDDEN MARKOV MODELS

BY KWEKU ABRAHAM[1,a], ELISABETH GASSIAT[2,b] AND ZACHARIE NAULET[2,c]

[1] *University of Cambridge, Statistical Laboratory, Wilberforce Road, Cambridge CB3 0WB, UK,* [a]*lkwa2@cam.ac.uk*

[2] *Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France,* [b]*elisabeth.gassiat@universite-paris-saclay.fr;* [c]*zacharie.naulet@universite-paris-saclay.fr*

Hidden Markov models (HMMs) are flexible tools for clustering dependent data coming from unknown populations, allowing nonparametric modelling of the population densities. Identifiability fails when the data is in fact independent, and we study the frontier between learnable and unlearnable two-state nonparametric HMMs. Interesting new phenomena emerge when the cluster distributions are modelled via density functions (the 'emission' densities) belonging to standard smoothness classes compared to the multinomial setting [2]. Notably, in contrast to the multinomial setting previously considered, the identification of a direction separating the two emission densities becomes a critical, and challenging, issue. Surprisingly, it is possible to "borrow strength" from estimators of the smoother density to improve estimation of the other. We conduct precise analysis of minimax rates, showing a transition depending on the relative smoothnesses of the emission densities.

1. Introduction. Consider a two-state HMM with real-valued emissions, in which we observe the first n entries of a sequence $\boldsymbol{Y} = (Y_1, Y_2, \dots) \in [0,1]^{\mathbb{N}}$ which, under a parameter $\theta = (p, q, f_0, f_1)$, satisfies

$$\begin{gathered}\mathbb{P}_\theta(Y_n \in A \mid \boldsymbol{X}) = \int_A f_{X_n}(y)\mathrm{d}y, \\ \boldsymbol{X} = (X_n)_{n\in\mathbb{N}} \sim \mathrm{Markov}(\pi, Q),\end{gathered} \tag{1}$$

with the $Y_n,\ n \in \mathbb{N}$ conditionally independent given $\boldsymbol{X}$. The vector $\boldsymbol{X}$ of 'hidden states', which we assume is started from its invariant distribution $X_1 \sim \pi$, takes values in $\{0,1\}^{\mathbb{N}}$. The transition matrix of the chain is given by

$$Q := \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, \tag{2}$$

with the convention that for $j \geq 1$, $\mathbb{P}_\theta(X_{j+1} = 0 \mid X_j = 0) = 1 - p < 1$ and $\mathbb{P}_\theta(X_{j+1} = 0 \mid X_j = 1) = q > 0$. The densities $f_0 \in B^{s_0}_{2,\infty}(R), f_1 \in B^{s_1}_{2,\infty}(R)$ are the 'emission densities' with respect to Lebesgue measure on $[0,1]$; here we use the notation

$$B^s_{2,\infty}(R) = \{f \in B^s_{2,\infty} : \|f\|_{B^s_{2,\infty}} \leq R\}$$

for the scaled unit ball of the Besov space $B^s_{2,\infty}$. The precise definition of $\|\cdot\|_{B^s_{2,\infty}}$ used in this paper is delayed to equation (10) below. We throughout use $\mathbb{P}_\theta$ to denote the law of $(\boldsymbol{X}, \boldsymbol{Y})$, and all induced marginal and conditional laws.

The goal is to estimate the parameter θ. This is known to be possible, up to a label-switching issue, under very mild conditions [10, 3]: specifically, given that the highly non-identifiable i.i.d. nonparametric mixture is a degenerate submodel of a HMM, under conditions which rule out independence. There are three ways in which the data $(Y_n)_{n\in\mathbb{N}}$ can

MSC2020 subject classifications: Primary 62M05, 62G05; secondary 62G07.
Keywords and phrases: Hidden Markov Models, Minimax, Nonparametric estimation.

fail to exhibit dependence: when the hidden states themselves are in reality independently distributed; when the emission distributions are identical; or when only one population is observed.

Once these i.i.d. submodels are excluded, consistent estimation is possible even for nonparametric emission distributions. Moreover, no cost is incurred relative to the case where the underlying labels are observed: for s-smooth functions, the minimax rate $n^{-s/(1+2s)}$ is achieved, see [8, 9]. This rate can be achieved adaptively in a "state-by-state" manner: up to a label-switching issue, one can achieve the rate $n^{-s_j/(1+2s_j)}$ if f_j has smoothness s_j, without knowledge of $(s_j,\ j=0,1)$, see [15]. See also [14] for robust estimation of the law of the observations in finite state space HMMs. These works do not consider the tradeoff between the required sample size and the required "distance" from independence, and it is this tradeoff that forms the focus of the current work, continuing from the previous article [2] in which we considered the model (1) but with f_0, f_1 multinomial densities with respect to counting measure on $\{1,\dots,K\}$ rather than the Lebesgue densities on $[0,1]$ as considered herein. The nonparametric setting exhibits striking *qualitative*, as well as *quantitative*, differences relative to the multinomial setting: see Section 1.2.

As in [2] we adopt the minimax paradigm and we analysis the smallest maximum risk attainable over the following class of parameters. We define for some $\delta,\epsilon\in(0,1)$ and some $\zeta, s_0, s_1, R>0$

$$\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R) \coloneqq \{\theta : p,q\geq\delta,\ |1-p-q|\geq\epsilon,\ \|f_0-f_1\|_{L^2}\geq\zeta,\ \|f_i\|_{B^{s_i}_{2,\infty}}\leq R\}. \tag{3}$$

The quantities δ,ϵ and ζ lower bound the "distance" to the i.i.d. submodel. Indeed if $\delta=0$, we may be unable to estimate both f_0 and f_1 since we may see data from one of these alone; if $\zeta=0$ we may be unable to estimate p and q; and if $\epsilon=0$ then we may be unable to identify the contributions of f_0 and f_1 to the mixture $\pi_0 f_0+\pi_1 f_1$. In contrast with [8, 9] which consider nonparametric estimation of f_0 and f_1 in the large δ,ϵ,ζ regime, we are mainly interested in the regimes where δ,ϵ,ζ can be eventually small, and how the minimax risks for Q and f_0, f_1 over $\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ are affected in these regimes.

The main message of our theorems may now be stated roughly as follows (up to label switching and technical details relative to smoothnesses).

- The minimax rate over $\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ for the estimation of the finite dimensional parameter Q is the same as in the multinomial (parametric) setting, $(n\delta^2\epsilon^4\zeta^6)^{-1/2}\max(\delta,\epsilon\zeta)$.
- The minimax rate over $\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ for the estimation of emission densities f_0 and f_1 exhibits a transition according to the relative smoothnesses of the densities. If $s_0 = s_1 = s$, then the minimax rate of estimating f_0 and f_1 in L^2 norm is $(\delta^2\epsilon^4\zeta^4 n)^{-1} + (n\delta^2\epsilon^2\zeta^2)^{-s/(2s+1)}$, while if $s_0>s_1$ (morally, see Section 4.4 for details), then the minimax rate for f_0 in L^2 norm is $(\delta^2\epsilon^4\zeta^4 n)^{-1} + (n\delta^2\epsilon^2\zeta^2)^{-s_0/(2s_0+1)}$ and the minimax rate for f_1 in L^2 norm is $(\delta^2\epsilon^4\zeta^4 n)^{-1} + (n\delta^2)^{-s_1/(2s_1+1)}$.
- There exist estimators achieving the minimax risk (up to constants) that are adaptive in the smoothness of the emission densities.

Suppressed constants may depend on R, on an upper bound L for the essential supremum of f_0 and f_1 and on a lower bound γ^* for the absolute spectral gap of the chain $\boldsymbol{X}$.

For full statements of the theorems see Section 3 (minimax rates for the estimation of Q), Sections 4.2 and 4.4 (upper bounds) and Section 4.3 (lower bounds). The precise theorems are stated in a nonasymptotic manner and upper bounds contain several terms with precise behaviour with respect to the constants R, L and γ^*. The asymptotic leading terms given in the above main results are in the case where the "distance" to frontier is large compared to n^{-a} for some (precisely defined) a. In this setting, the transition between the situation where emission densities have similar or different smoothnesses can be described as "$s_0 = s_1$" or "$s_0 > s_1$", but the transition appears in a more intricate manner when taking a nonasymptotic point of view. The exact relationship between s_0 and s_1 required is described in Section 4.4. However, the main message is that some transition in the minimax rate occurs depending on the relative smoothnesses of the emission densities.

1.1. *Comparison to the multinomial case.* Let us compare the above theorem to what was obtained in the multinomial case in [2]. There, in place of the Besov classes $B^{s_0}_{2,\infty}(R)$, $B^{s_1}_{2,\infty}(R)$, we considered density vectors, i.e. densities with respect to counting measure on $\{1, \dots, K\}$ for some $K \geq 2$. An identifiability assumption that $f_1 - f_0$ lies in some specified half plane was taken, to avoid any label switching issues.

The main result of that work can be summarised as follows. The worst-case (minimax) risk for estimating parameters are the following (suppressed constants may depend on K and on γ^*):

- The transition matrix Q can be estimated at minimax rate $(n\delta^2\epsilon^4\zeta^6)^{-1/2}\max(\delta, \epsilon\zeta)$;
- The multinomial density vectors can each be estimated at minimax rate $(n\delta^2\epsilon^4\zeta^4)^{-1/2}$.

We note that the parametric part $\hat{Q}$ achieves the same rate in the nonparametric setting as in the multinomial setting; at first glance this seems unsurprising in view of the fact that the pairs $((X_n, h(Y_n))_{n\geq 0}$ form a hidden Markov model with transition matrix Q for any function h, so that for a suitable h we can reduce to a parametric setting. However, reducing to a parametric setting in which Q is still identifiable is in fact a nonparametric problem, see Section 2.3, so that getting the same minimax parametric rate is not a priori guaranteed.

In contrast the rates for f_0 and f_1 in the nonparametric setting arise from delicate interplay between the smoothnesses s_0, s_1 and the parameters δ, ϵ, ζ appear with different powers.

1.2. *Novelties in the nonparametric setting.* One typical approach in nonparametric statistics is to reduce to a multinomial case by projecting onto a wavelet basis with maximum level chosen to balance the bias and variance. This basic approach broadly works here but the inverse nature of the problem, specifically with regards to inverting the map $\theta \mapsto \mathbb{P}_\theta$, introduces some novelties.

Separating the two distributions. In the HMM setting, the key issue of separation from the independent subcase becomes entangled with the choice of a direction in which to project f_0 and f_1. Let us illustrate this point in the context of estimating the parametric part Q. As noted above, for any function h the pairs $((X_n, h(Y_n))_{n\geq 0}$ form a hidden Markov model with transition matrix Q. This is the *no bias* phenomenon already used in [11] for multidimensional mixture models and in [16] for finite state space HMMs. Choosing $A_1, \dots, A_K$ partitioning $\mathcal{Y}$ and defining h by $h(y) = k$ for $y \in A_k$, we may apply the results from the multinomial setting to deduce that Q can be estimated at the rate given in Section 1.1. However in said rate ζ must lower bound the euclidean distance between vectors $(\langle f_0, \mathbb{1}_{A_k}\rangle : k \leq K)$ and $(\langle f_1, \mathbb{1}_{A_k}\rangle : k \leq K)$. If the A_k are not chosen carefully, this distance may be much smaller than $\|f_0 - f_1\|_{L^2}$, potentially even equal to 0. The suitable choice of A_k depends on the direction $(f_0 - f_1)/\|f_0 - f_1\|_{L^2}$, which is unknown and *nonparametric*. One must therefore account for the nonparametric modelling of the emission distributions even in estimating the parametric portion. The requirement on an initial estimator is not too stringent: one sufficient condition, which we will use in what follows, is to find h of norm 1 such that $\langle h, f_0 - f_1\rangle/\|f_0 - f_1\|_{L^2}$ is bounded away from zero. Finding such an h is nevertheless an interesting challenge in the current setting, since we consider the case where f_0 and f_1 are potentially very close. This preliminary estimation step is important also for estimating the nonparametric part.

Sharing estimation strength. Another novelty in the nonparametric setting is a "coupling" that appears between estimation for f_0 and estimation for f_1. Indeed, it is possible to estimate the combination $\psi_1 = \pi_0 f_0 + \pi_1 f_1$ at a fast rate because we avoid the inversion step: ψ_1 is simply the invariant density of Y_n, and so an empirical estimator achieves the nonparametric rate $n^{-s/(1+2s)}$ where s is the smoothness of ψ_1 (this can be proven using Lemma 7). In the case where f_0 is much smoother than f_1, it may be more efficient to estimate f_0 and ψ_1, and estimate f_1 by plug in, rather than directly estimating f_1. This is reflected both in the upper bounds (see Theorem 3 and Theorem 5) and the lower bounds (see Theorem 4). The precise analysis of how one can "borrow" strength from the estimator of the smoother emission density to improve on the estimation rate for the rougher emission density is more involved, but this is the ground idea.

Choice of wavelet thresholding estimator. One additional novelty relative to other HMM papers in the nonparametric setting is that we use a wavelet block thresholding estimator. This allows us to adapt to the smoothnesses s_0 and s_1 without needing to use Lepskii's method, and so in principle at least is more computationally feasible. One could also achieve adaptive rates using updated versions of previous, Lepskii's method based, estimators, from [15].

1.3. *Outline of the paper.* In Section 2, we recall the reparametrisation trick that illuminates how the distance to the i.i.d. frontier appears to be key for being able to solve the inverse problem, together with the need to find a way to separate the two emission distributions. In Section 3 we focus first on the estimation of the transition matrix Q, for which we provide a new moment-based estimator and precise minimax rates, see Theorem 1 for the upper bound and Theorem 2 for the lower bound. Section 4 studies estimation of the emission densities with $L^2[0,1]$-norm risk. Using the strategy of Section 2 to solve the inverse problem, we provide a new block-thresholding wavelet based estimation method for which we give in Theorem 3 precise upper bounds for the maximum risk, adaptively achieving the usual asymptotic minimax rate. Lower bounds are proved in Theorem 4 for which a transition depending on the relative smoothnesses of the emission distributions appears. A consequence is that for similar smoothnesses the previous estimator achieves the asymptotic minimax rate even with respect to constants governing the distance to the frontier as soon as this distance is not too small compared to some negative power of n. In Section 4.4 we propose another estimator which can handle the transition if we can identify which is the smoother function, see Theorem 5. Section 5 explains how to separate the emission distributions, see Theorem 6. Section 6 is devoted to the discussion of questions left open in our work, in particular the question of full adaptation to distance to the frontier with respect to unknown possible different smoothnesses in the emission densities. All detailed proofs are given in Section A (upper bounds) and Section B (lower bounds).

2. Key elements for solving the inverse problem.

2.1. *Reparametrisation.* As noted previously [10], understanding law of three consecutive observations is key to solving the inverse problem and recovering the model parameters. As in the multinomial case, a reparametrisation simplifies the expression for said law, and allows the dependence on the parameters δ, ϵ and ζ to appear more naturally.

Set

$$\phi(\theta) = \left(\tfrac{q-p}{p+q}\ 1-p-q\ \|f_0 - f_1\|_{L^2}\right), \quad \psi(\theta) = \left(\tfrac{qf_0+pf_1}{p+q}\ \tfrac{f_0-f_1}{\|f_0-f_1\|_{L^2}}\right). \tag{4}$$

For $m \geq 1$, let $P^{(m)}_{\phi,\psi}$ denote the law of $(Y_1, \ldots, Y_m)$ under parameter (ϕ, ψ), and let $p^{(m)}_{\phi,\psi}$ denote the corresponding density with respect to Lebesgue measure on $[0,1]^m$. In this parametrisation, defining

$$r(\phi) = \tfrac{1}{4}(1-\phi_1^2)\phi_2\phi_3^2, \tag{5}$$

one computes, with $f \otimes g$ defined by $(f \otimes g)(x,y) = f(x)g(y)$,

$$\begin{aligned} p^{(3)}_{\phi,\psi} = \psi_1 \otimes \psi_1 \otimes \psi_1 + r(\phi)\big(\psi_2 \otimes \psi_2 \otimes \psi_1 + \psi_1 \otimes \psi_2 \otimes \psi_2\big) \\ + \phi_2 r(\phi)\psi_2 \otimes \psi_1 \otimes \psi_2 - \phi_1\phi_2\phi_3 r(\phi)\psi_2 \otimes \psi_2 \otimes \psi_2. \end{aligned} \tag{6}$$

The parametrisation $\theta \mapsto (\phi, \psi)$ is invertible: see Lemma 1. It is also possible to invert the map $(\phi, \psi) \mapsto p^{(3)}_{\phi,\psi}$ up to label switching issues.

2.2. *Solving the direct problem.* Of course, to solve the inverse problem we must solve the direct problem (which here means estimating $p^{(3)}_{\phi,\psi}$ or equivalent). In [2], where multinomial emission densities were considered, it was proposed to use an empirical estimator $\hat{p}^{(3)}$ of $p^{(3)}$, and then solve the inverse problem using an estimator $(\hat{\phi}, \hat{\psi})$ minimizing the (euclidean) distance $(\phi, \psi) \mapsto \|p^{(3)}_{\phi,\psi} - \hat{p}^{(3)}\|$. Here we use a similar heuristic, but we propose using the method of moments to improve tractability. As explained in Section 1.2, we need to have access to a separating hyperplane, or equivalently the unit normal to such a plane, which will be a "sufficiently good" separating function $\tilde{\psi}_2$ (preliminary estimator of ψ_2), discussed in the next subsection. Given $\tilde{\psi}_2$, by considering the expectations of $\tilde{\psi}_2$, $\tilde{\psi}_2 \otimes \tilde{\psi}_2$, $\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2$ and $\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2$, from (6) one can extract

$$m = m(\phi) := (r(\phi)\tilde{\mathcal{I}}^2, r(\phi)\phi_2\tilde{\mathcal{I}}^2, r(\phi)\phi_1\phi_2\phi_3\tilde{\mathcal{I}}^3), \quad \tilde{\mathcal{I}} := \langle \psi_2, \tilde{\psi}_2 \rangle; \tag{7}$$

see Lemma 2. If $\tilde{\mathcal{I}} \neq 0$, one can retrieve from m the parametric part (i.e. the parameters (p,q), or equivalently (ϕ_1, ϕ_2)): for example, $\phi_2 = m_2/m_1$. We therefore propose a method of moments estimator for (p,q) constructed via estimating the expectations above using their empirical versions, see Section 3, computed using

$$\mathbb{P}^{(s)}_n(h) := \frac{1}{n-s+1}\sum_{i=1}^{n-s+1} h(Y_i, \ldots, Y_{i+s-1}), \quad h : [0,1]^s \to \mathbb{R},\ s \geq 1. \tag{8}$$

We use concentration inequalities for Markov chains [17] to ensure our empirical estimators are sufficiently close to their means. This requires us to slightly shrink the set $\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ and restrict our attention to parameters that are also in (see Lemma 7)

$$\Sigma_{\gamma^*}(L) := \{\theta : 1 - |1-p-q| \geq \gamma^*,\ \max_{j=0,1}\|f_j\|_\infty \leq L\}, \tag{9}$$

i.e. parameters with uniformly bounded emission densities (here $\|\cdot\|_\infty$ denotes the usual supremum norm) and having an absolute spectral gap. Note that for $\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ we automatically have $\|f_j\|_\infty \lesssim R$, but we introduce L to highlight the distinct role played by $\|f_j\|_\infty$ from that played by R. Also note that we distinguish between the parameters δ, ϵ, ζ which relate to the "distance" to independence and the others.

To estimate the nonparametric part, i.e. f_0 and f_1, we use empirical estimators of their respective wavelet coefficients based on $\mathbb{P}^{(s)}_n(h)$ for well-chosen h, using again the separating function $\tilde{\psi}_2$, see Section 4. Inversion to get the emission densities again requires that $\tilde{\mathcal{I}} \neq 0$.

2.3. *Nonparametric estimation of a separating hyperplane.* As described above, provided $\tilde{\mathcal{I}} = \langle\tilde{\psi}_2, \psi_2\rangle \neq 0$ one can build a map

$$\big(\mathbb{E}_\theta(\tilde{\psi}_2), \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2), \mathbb{E}_\theta(\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2), \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2)\big) \mapsto m(\phi) \mapsto (p, q),$$

up to label switching (and similarly to recover the nonparametric part). Then an estimator $(\hat{p}, \hat{q})$ of (p, q) is obtained via the method of moments by evaluating the previous map at $(\mathbb{P}_n^{(1)}(\tilde{\psi}_2), \mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2), \mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2), \mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2))$. When $\tilde{\mathcal{I}}$ is too small, however, said map turns out to be unstable and $(\hat{p}, \hat{q})$ may end up being far from (p, q) even though the empirical moments are close to their expectations. Hence one requires $|\tilde{\mathcal{I}}|$ to be bounded away from zero to avoid deteriorated rates of convergence. Although seemingly innocuous, this has great consequences when working close to the i.i.d. boundary. To illustrate this claim, suppose that we choose $\tilde{\psi}_2$ randomly as follows (other random mechanisms will lead to similar issues): let $(e_k)_{k\geq 1}$ be an orthonormal basis for $L^2[0,1]$, for $K \geq 1$ draw $U_1, \ldots, U_K \overset{iid}{\sim} \mathcal{N}(0,1)$, and let $\tilde{\psi}_2 = \sum_{k=1}^K U_k e_k / [\sum_{m=1}^K U_k^2]^{1/2}$. Then, for K large enough,

$$\langle\tilde{\psi}_2, \psi_2\rangle = \frac{\sum_{k=1}^K U_k \langle e_k, f_0 - f_1\rangle}{[\sum_{m=1}^K U_k^2]^{1/2} \|f_0 - f_1\|_{L^2}} \overset{\text{law}}{\approx} \mathcal{N}(0,1) \cdot \frac{\|\pi_K f_0 - \pi_K f_1\|_{L^2}}{\sqrt{K}\|f_0 - f_1\|_{L^2}}$$

with $\pi_K f_0 - \pi_K f_1$ the orthogonal projection of $f_0 - f_1$ onto the span of $e_1, \ldots, e_K$. In a maximum risk analysis we cannot exclude that $\|\pi_K f_0 - \pi_K f_1\|_{L^2} \ll \|f_0 - f_1\|_{L^2}$ unless K is taken very large or going to infinity fast enough. But then, $\langle\tilde{\psi}_2, \psi_2\rangle$ will be $O_p(K^{-1/2})$ and harm the rates of convergence of our estimators. This simple example shows that $\tilde{\psi}_2$ must be determined from a priori knowledge or computed from the data. We note that this is an issue only when ζ is very small and/or $f_0 - f_1$ is non-smooth. Otherwise, when working far from the i.i.d. boundary, choosing a random $\tilde{\psi}_2$ may work reasonably well. Indeed, similar heuristics have been used in other papers with good performances [1, 9, 16].

Thus a main step to be able to use (6) to solve the inverse problem is to build a function $\tilde{\psi}_2$ such that $|\tilde{\mathcal{I}}|$ is bounded away from zero. This explains why we describe $\tilde{\psi}_2$ as a separating function: since $\psi_2 \coloneqq (f_0 - f_1)/\|f_0 - f_1\|_{L^2}$, finding $\tilde{\psi}_2$ is tantamount to finding an hyperplane in $L^2[0,1]$ which separates f_0 and f_1 sufficiently well. This means that even for estimating the parametric part, we must first solve for the nonparametric problem of finding $\tilde{\psi}_2$. As explained in the introduction, this step is an important difference to the multinomial situation where densities lived in a finite dimensional space, and it is therefore not trivial that (p, q) can be estimated at parametric speed in the current setting.

Since $f \mapsto |\langle\psi_2, f\rangle|$ is maximized over the unit ball when $f = \psi_2$, the best choice for $\tilde{\psi}_2$ is an estimator of ψ_2. Since we only require that $|\tilde{\mathcal{I}}| = |\langle\psi_2, \tilde{\psi}_2\rangle|$ bounded away from zero, said estimator does not necessarily need to be good. One way to build such an estimator is to start with a truncated orthonormal basis $(e_k)_{k=1,\ldots,K}$ for $L^2[0,1]$ and define the $K \times K$ matrix $\mathcal{G}$ with entries

$$\mathcal{G}_{jk} \coloneqq \frac{1}{2}\mathbb{E}_\theta(e_j \otimes e_k + e_k \otimes e_j) - \mathbb{E}_\theta(e_j)\mathbb{E}_\theta(e_k) = r(\phi)\langle\psi_2, e_j\rangle\langle\psi_2, e_k\rangle$$

where the last equality follows from Lemma 2. Hence, $\mathcal{G}$ is proportional to the Gram matrix of the vector $V_\theta \propto (\langle\psi_2, e_1\rangle, \ldots, \langle\psi_2, e_K\rangle)$. Clearly $V_\theta/\|V_\theta\|$ is the sole eigenvector of $\mathcal{G}$ with a non-zero eigenvalue, this unique non-zero eigenvalue being equal to $r(\phi)\sum_{k=1}^K \langle\psi_2, e_k\rangle^2$. Hence, the basis coefficients of ψ_2 can be recovered from $\mathcal{G}$. By concentration arguments, we expect that an estimator of ψ_2 can be obtained from the empirical version of $\mathcal{G}$. We leverage that idea using a wavelet basis in Section 5.

Since we ultimately wish to compute empirical averages of the form $\mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes f)$, it is convenient to assume that $\tilde{\psi}_2$ is independent of the data Y, for example because it expresses a priori knowledge or because it is computed using the above method from a sample

$(\tilde{Y}_1, \dots, \tilde{Y}_n) \sim P_\theta^{(n)}$ that is independent of $(Y_1, \dots, Y_n)$. This will be always assumed in the sequel and we come back to discuss this point in Section 5.

3. Estimation of the parametric part and minimax rates. First we estimate the m functional defined in (7), using the method of moments. Drawing inspiration from expressions in Lemma 2, we let

$$\begin{aligned}
\hat{m}_1 &:= \mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2) - \mathbb{P}_n^{(1)}(\tilde{\psi}_2)^2, \\
\hat{m}_2 &:= \mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2) - \mathbb{P}_n^{(1)}(\tilde{\psi}_2)^2, \\
\hat{m}_3 &= -\mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2) + \mathbb{P}_n^{(1)}(\tilde{\psi}_2)^3 + \big(2\hat{m}_1 + \hat{m}_2\big)\mathbb{P}_n^{(1)}(\tilde{\psi}_2),
\end{aligned}$$

and then

$$\hat{\phi}_1 = \frac{\hat{m}_3}{[4\hat{m}_1^2(\hat{m}_2)_+ + \hat{m}_3^2]^{1/2}}, \qquad \hat{\phi}_2 = \max\Big(-1, \min\Big(\frac{\hat{m}_2}{\hat{m}_1}, 1\Big)\Big).$$

We then build an estimator of the transition matrix Q, justified by Lemma 1, by letting

$$\begin{aligned}
\hat{Q}_{01} &= 1 - \hat{Q}_{00} = \tfrac{1}{2}(1 - \hat{\phi}_1)(1 - \hat{\phi}_2), \\
\hat{Q}_{10} &= 1 - \hat{Q}_{11} = \tfrac{1}{2}(1 + \hat{\phi}_1)(1 - \hat{\phi}_2).
\end{aligned}$$

To account for label switching, write Q_σ for the matrix with entries $(Q_\sigma)_{ij} = Q_{\sigma(i),\sigma(j)}$ for a permutation σ. We consider the loss relative to the Frobenius norm $\|\cdot\|_{\mathrm{F}} := \sum_{i,j}(\cdot)_{i,j}^2$.

THEOREM 1. *Assume $\zeta \le 1$, and assume $\tilde{\psi}_2$ is a given unit vector in $L^2[0,1]$, independent of the sample $Y_1, \dots, Y_n$, satisfying $\|\tilde{\psi}_2\|_\infty \le \tau$ and $|\langle \tilde{\psi}_2, \psi_2 \rangle| \ge 7/8$. Assume $n\gamma^* \ge \tau^6/L^3$. Then there are universal constants $B, C > 0$ such that*

$$\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} \inf_\sigma \mathbb{E}_\theta\Big(\|\hat{Q}_\sigma - Q\|_{\mathrm{F}}^2\Big)$$
$$\le B \exp\left(-\frac{Cn\gamma^*\delta^2\epsilon^4\zeta^6}{L^3 + \max(\tau, \sqrt{L})^3\delta\epsilon^2\zeta^3}\right) + \frac{BL^3\max(\delta^2, \epsilon^2\zeta^2)}{\delta^2\epsilon^4\zeta^6}\frac{1}{n\gamma^*}.$$

We assumed that $|\langle \tilde{\psi}_2, \psi_2 \rangle| \ge 7/8$ in the statement of the Theorem 1. As discussed earlier in Section 2.3 what matters the most is that $|\langle \tilde{\psi}_2, \psi_2 \rangle|$ is bounded away from zero. The lower bound of $7/8$ is somewhat arbitrary and inspired from the further results in Section 5. Having a smaller value for $|\langle \tilde{\psi}_2, \psi_2 \rangle|$ would only affect the upper bound in Theorem 1 through the value of the constant B. It is also required in Theorem 1 that $\|\tilde{\psi}_2\|_\infty \le \tau$ for some $\tau > 0$ [note that $\|\tilde{\psi}_2\|_{L^2} = 1$, so necessarily $\tau \ge 1$] which comes from technicalities arising in the proof from the use of certain Bernstein-type concentration inequalities. The same remarks will also apply to the subsequent Theorems 3 and 5.

In most regimes of interest the first term in the bound in Theorem 1 can be neglected and our estimator achieves the rate of convergence $\frac{L^3\max(\delta^2,\epsilon^2\zeta^2)}{\delta^2\epsilon^4\zeta^6}\frac{1}{n\gamma^*}$, which is, up to constants, the minimax rate, as the lower bound below proves.

We give a lower bound for each component of Q separately, which obviously implies a bound for the left side of the above display.

THEOREM 2. *Assume* $n\delta^2\epsilon^4\zeta^6 \geq 1$, $\zeta \leq 1/(4\sqrt{3})$, $\gamma^* \leq 1/3$, $\epsilon \leq \epsilon_0$ *for a suitable* $\epsilon_0 > 0$, $\delta \leq 1/6$, $R \geq 5/4 + 1/(8\sqrt{3})$ *and* $L \geq 5/8$. *Then there exists a constant* $c > 0$ *such that*

$$\inf_{\hat\theta} \sup_{\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta\big(|\hat p - p|^2\big) \geq \frac{c \max(\delta^2, \epsilon^2\zeta^2)}{\delta^2\epsilon^4\zeta^6}\frac{1}{n}$$

where the infimum is over all estimators $\hat\theta$ *based on* $Y_1, \ldots, Y_n$. *The same lower bound holds for the estimation of* q.

The proof of Theorem 1 can be found in Section A.3 and that of Theorem 2 in Section B.1.

4. Estimation of the emission densities.

4.1. *Preliminaries on wavelets and the space* $B^s_{2,\infty}$. Throughout the paper we use the S-regular boundary-corrected wavelet basis of [7], see also e.g. [12, Section 4.3.5], denoted $\{\{\Phi_{Jk} : k = 0, \ldots, 2^{J-1}\}, \{\Psi_{jk} : j \geq J, k = 0, \ldots, 2^j - 1\}\}$, with initial resolution level J chosen as in the latter reference. As is common, we will refer to the (Φ_{Jk}) as father wavelets and to the (Ψ_{jk}) as mother wavelets. Any $f \in L^2[0,1]$ has the series expansion

$$f = \sum_{k=0}^{2^J-1} \langle \Phi_{Jk}, f\rangle \Phi_{Jk} + \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} \langle \Psi_{jk}, f\rangle \Psi_{jk}$$

with convergence of the series in $L^2[0,1]$. In fact, as our densities will be assumed regular enough, wavelet series expansions for f_0 and f_1 will also converge uniformly (see [12, eq. (4.71)]). Furthermore, it is well-known that the Besov space $B^s_{2,\infty}$ can be characterised via the wavelet coefficients. Indeed the norm for $B^s_{2,\infty}$ that we will use (see e.g. [12, Equation (4.166)]) is given by

$$\|f\|^2_{B^s_{2,\infty}} := \sum_{k=0}^{2^J-1} \langle \Phi_{Jk}, f\rangle^2 + \sup_{j \geq J} 2^{2js} \sum_{k=0}^{2^j-1} \langle \Psi_{jk}, f\rangle^2. \tag{10}$$

4.2. *Block wavelet estimators achieve smoothness adaptive rates.* Using the ideas in Section 2, the coefficients of f_0 and f_1 can be extracted from $\{\mathbb{E}_\theta(\tilde\psi_2 \otimes \Phi_{Jk})\}$, $\{\mathbb{E}_\theta(\tilde\psi_2 \otimes \Psi_{jk})\}$, $\{\mathbb{E}_\theta(\Phi_{Jk})\}$, $\{\mathbb{E}_\theta(\Psi_{jk})\}$ and $\mathbb{E}_\theta(\tilde\psi_2)$, and further estimated using their empirical relatives. Given these empirical wavelets coefficients, we construct estimators for f_0 and f_1 based on block-thresholding the coefficients.

For notational convenience, we write $f^{\Phi_{Jk}} := \langle \Phi_{Jk}, f\rangle$ and $f^{\Psi_{jk}} := \langle \Psi_{jk}, f\rangle$. Hence, our goal is to find estimators $\{(\hat f_0^{\Phi_{Jk}})_k, (\hat f_0^{\Psi_{jk}})_{jk}\}$ of $\{(f_0^{\Phi_{Jk}})_k, (f_0^{\Psi_{jk}})_{jk}\}$ (and similarly for f_1). To obtain an expression for these coefficients, we draw inspiration from the inversion formulae in Lemma 1. In particular, letting

$$\hat g := \frac{\sqrt{4\hat m_1^2(\hat m_2)_+ + \hat m_3^2}}{\hat m_2}\mathbf{1}_{\{\hat m_2 > 0\}}$$

be an estimator of $g := \phi_3|\tilde{\mathcal{I}}|$, we set

$$\hat G^{\Phi_{Jk}} := \mathbb{P}_n^{(2)}(\tilde\psi_2 \otimes \Phi_{Jk}) - \mathbb{P}_n^{(1)}(\tilde\psi_2)\mathbb{P}_n^{(1)}(\Phi_{Jk}),$$

$$\hat f_0^{\Phi_{Jk}} := \mathbb{P}_n^{(1)}(\Phi_{Jk}) + \frac{\hat g(1-\hat\phi_1)}{2\hat m_1}\mathbf{1}_{\{\hat m_1 \neq 0\}}\hat G^{\Phi_{Jk}},$$

$$\hat f_1^{\Phi_{Jk}} := \mathbb{P}_n^{(1)}(\Phi_{Jk}) + \frac{\hat g(1+\hat\phi_1)}{2\hat m_1}\mathbf{1}_{\{\hat m_1 \neq 0\}}\hat G^{\Phi_{Jk}}.$$

Lemmas 1 and 2 (and the sentence after the latter) justify that these target coefficients of f_0, f_1 (up to label switching). The same definition applies *mutatis mutandis* to the estimators of the mother coefficients $\hat{f}_0^{\Psi_{jk}}$, $\hat{f}_1^{\Psi_{jk}}$, and $\hat{G}^{\Psi_{jk}}$. It is customary that not all empirical coefficients be retained in the final estimator, and that small coefficients should be discarded to reduce the risk. It is also well-known [5] that individual coefficient thresholding is sub-optimal with respect to the L^2 loss, as opposed to block-thresholding procedures with carefully chosen blocks [4, 6]. Here, we build the blocks as follows.

Motivated by [4, 6] we wish to build blocks of consecutive wavelets with size approximately $\log(n)$, which is known to be the best compromise for global versus local adaptation. Since there might be fewer than $\log(n)$ wavelets at small resolution level j, we will only threshold coefficients with j large enough. We define

$$J_n \coloneqq \inf\{j \geq J : 2^j \geq \log(n)\}$$

where the infimum is over the integers. We then let $N \coloneqq 2^{J_n}$ so that each level with $j \geq J_n$ can be partitioned into an integer number of blocks of N consecutive wavelets. More precisely, for each level $j \geq J_n$, and each $\ell = 0, \ldots, N^{-1}2^j - 1$ we define the blocks of indices

$$\mathfrak{B}_{j\ell} \coloneqq \{k \in \{0, \ldots, 2^{j-1}\} : (\ell - 1)N \leq k \leq \ell N - 1\}. \tag{11}$$

For a constant $\tau \geq 1$ we also define $\tilde{\jmath}_n$ as the largest integer such that $2^{\tilde{\jmath}_n} \leq \frac{n}{\log(n)\tau^2}$; we shall assume that $J < J_n < \tilde{\jmath}_n$ which is always satisfied for n large enough. We then let, for $i = 0, 1$,

$$\hat{f}_i \coloneqq \sum_{k=0}^{2^J-1} \hat{f}_i^{\Phi_{Jk}} \Phi_{Jk} + \sum_{j=J}^{J_n-1} \sum_{k=0}^{2^j-1} \hat{f}_i^{\Psi_{jk}} \Psi_{jk} + \sum_{j=J_n}^{\tilde{\jmath}_n} \sum_{\ell} \Bigg(\sum_{k \in \mathfrak{B}_{j\ell}} \hat{f}_i^{\Psi_{jk}} \Psi_{jk} \Bigg) \mathbf{1}_{\{\|\hat{f}_i^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{S}_n\}}$$

where $\|\hat{f}_i^{\mathfrak{B}_{j\ell}}\|^2 \coloneqq \sum_{k \in \mathfrak{B}_{j\ell}} (\hat{f}_{\pm}^{\Psi_{jk}})^2$, $\Gamma > 0$ is a tuning parameter, and

$$\hat{S}_n \coloneqq \sqrt{\frac{\log(n)}{n}} \max\Big(1, \frac{\hat{g}}{|\hat{m}_1|}\Big) \mathbf{1}_{\{\hat{m}_1 \neq 0\}}.$$

The above estimators perform well in probability; to ensure good perfomance in expectation we truncate below at 0 and above at some $\check{T}$, defining for $i = 0, 1$

$$\check{f}_i \coloneqq \max\big(0, \min\big(\check{T}, \hat{f}_i\big)\big).$$

THEOREM 3. *Suppose $n\gamma^* \geq \max(\tau^3, \frac{\tau^2 \log(n)^2}{L})$, $\tilde{\jmath}_n > J_n$, $L \leq n$, $\tau \geq 1$, $\check{T} \geq L$, and $\zeta \leq 1$. Assume $\tilde{\psi}_2$ is a given unit vector in $L^2[0,1]$, independent of the sample $Y_1, \ldots, Y_n$, satisfying $\|\tilde{\psi}_2\|_\infty \leq \tau$ and $|\langle \tilde{\psi}_2, \psi_2 \rangle| \geq 7/8$. Then there are universal constants $\beta > 0$, $B > 0$ and $C > 0$ such that for all $\Gamma \geq \beta L^{1/2} \max((L/\gamma^*)^{1/2}, 1/\gamma^*)$ and for $i = 0, 1$, provided $0 < s_i \leq S$ with $S > 0$ the regularity of the wavelet basis,*

$$\begin{aligned}
\sup_{\theta \in \Theta^{s_0, s_1}_{\delta, \epsilon, \zeta}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta \min_{i'=0,1} \Big(\|\check{f}_{i'} - f_i\|_{L^2}^2 \Big) &\leq B\check{T}^2 \exp\Bigg(- \frac{Cn\gamma^* \delta^2 \epsilon^4 \zeta^6}{L^3 + \max(\tau, \sqrt{L})^3 \delta \epsilon^2 \zeta^3} \Bigg) \\
&+ \frac{BL^2}{\delta^2 \epsilon^2 \zeta^2} \frac{\log(n)}{n\gamma^*} + \frac{BL^3}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{n\gamma^*} + \frac{B \max(\tau, \sqrt{L})^6}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{(n\gamma^*)^2} \\
&+ \frac{BR^2 \max(1, \frac{L^2}{\Gamma^2 \gamma^*})}{\min(1, s_i)} \Big(\frac{\Gamma^2}{R^2 \delta^2 \epsilon^2 \zeta^2 n} \Big)^{2s_i/(2s_i+1)} + \frac{BR^2 \max(1, \frac{L^2}{\Gamma^2 \gamma^*})}{\min(1, s_i)} \Big(\frac{\tau^2 \log(n)}{n} \Big)^{2s_i}.
\end{aligned}$$

The proof of Theorem 3 is in Section A.4. Of particular interest is the boundary regime, where γ^*, R, L, $\check{T}$ and τ are of constant order while δ, γ and ζ are small. The following corollary is intended to illustrate how the bound simplifies in such setting. The proof of Corollary 1 is given in Section A.7.

COROLLARY 1. *Assume that γ^*, R, L, $\check{T}$, and τ remain constant as $n \to \infty$ and $\delta \geq n^{-a}$, $\epsilon \geq n^{-b}$, $1 \geq \zeta \geq n^{-c}$ for constants $a, b, c > 0$ such that $1 - 2a - 4b - 6c > 0$ (this latter requirement corresponds to where the bounds on the right vanish, so that parameters are proved to be learnable). Then the bound in the Theorem 3 simplifies: for large enough n,*

$$\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta \min_{i'=0,1} \left(\|\check{f}_{i'} - f_i\|_{L^2}^2 \right) \leq C \left\{ \frac{1}{\delta^2 \epsilon^4 \zeta^4 n} + \left(\frac{1}{\delta^2 \epsilon^2 \zeta^2 n} \right)^{2s_i/(1+2s_i)} \right\},$$

for a constant C depending on γ^*, L, R, Γ, B, τ, $\check{T}$, and a, b, c.

4.3. *Lower bounds.* The following theorem gives lower bounds for the estimation risk of the emission densities. The detailed proof can be found in Section B.2.

THEOREM 4. *Assume $n\delta^2\epsilon^2\zeta^4 \geq 1$, $\zeta \leq 1/(4\sqrt{3})$, $\gamma^* \leq 1/3$, $\epsilon \leq \epsilon_0$ for a suitable $\epsilon_0 > 0$, $\delta \leq 1/6$, $R \geq 5/4 + 1/(8\sqrt{3})$ and $L \geq 5/8$. Then there exists a constant $c > 0$ such that*

$$\inf_{\check{f}_0} \sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta \left(\|\check{f}_0 - f_0\|_{L^2}^2 \right) \geq c \left\{ \frac{1}{\delta^2 \epsilon^4 \zeta^4 n} + \left(\frac{1}{\delta^2 n} \right)^{2s_0/(2s_0+1)} \right\}. \tag{12}$$

If moreover for suitable constants c_0 and c_1, it holds $(n\delta^2\epsilon^2\zeta^4)^{-s_0/(1+2s_0)} \leq c_0\zeta$ and $\delta^{2s_1+1}(n\epsilon^2\zeta^2)^{(s_1-s_0)} \leq c_1$, then there exists a constant $c > 0$ such that

$$\inf_{\check{f}_0} \sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta \left(\|\check{f}_0 - f_0\|_{L^2}^2 \right) \geq c \left\{ \frac{1}{\delta^2 \epsilon^4 \zeta^4 n} + \left(\frac{1}{\delta^2 \epsilon^2 \zeta^2 n} \right)^{2s_0/(2s_0+1)} \right\}. \tag{13}$$

The infima are over all estimators $\check{f}_0$ based on $Y_1, \ldots, Y_n$. The same lower bounds hold for the estimation of f_1 by exchanging the role of s_0 and s_1 in the conditions and in the bounds.

This theorem calls for a number of comments. In the particular situation where $s_0 = s_1$, the lower bound (13) holds for the estimation of both emission densities, and the estimator described in Section 4.2 is rate minimax adaptive, including to the parameters of interest δ, ϵ, ζ.

The first part of the theorem states that for the estimation of the emission densities, the minimax risk is lower bounded by a parametric term which is similar to the one obtained in the multinomial situation, and a nonparametric term with the usual rate $n^{-2s_0/(2s_0+1)}$ corrected with δ^2, that is with an effective sample size $\delta^2 n$ replacing n. This shows that the inverse problem fundamentally makes estimation harder: if we were to observe X, we would on average see $n\pi_0 \gtrsim n\delta$ i.i.d. samples from f_0, hence would be able to estimate this with effective sample size $n\delta$, which may be much larger than $n\delta^2$.

The second part of the theorem is more involved. It states that, if one of the emission density is smooth enough compared to the other one and relatively to frontiers parameters, the lower bound can be made larger, with an effective sample size $\delta^2\epsilon^2\zeta^2 n$. This occurs for instance when $s_0 \geq s_1$. Thus, the smoothest emission density gets the smallest effective sample size when getting close to the frontier.

Now, the question remains: if indeed one of the emission density is much smoother than the other, can we improve the estimation of the one that is less smooth so that the upper bound for the maximum risk matches, up to constants, the lower bound (12)? In the next section, we propose an estimation procedure proving that the lower bound (12) is indeed sharp, and we discuss adaptation.

4.4. *Matching the upper and the lower bounds.* In this section, we propose another estimation procedure for the emission densities, with the aim of sharing the estimation strength of the smoother one with the rougher one. The starting point is to remark that

$$f_0 = \frac{2\psi_1}{1+\phi_1} - \left(\frac{1-\phi_1}{1+\phi_1}\psi_1 - \frac{g(1-\phi_1)}{2m_1}G\right) \text{ and } f_1 = \frac{2\psi_1}{1-\phi_1} - \left(\frac{1+\phi_1}{1-\phi_1}\psi_1 + \frac{g(1+\phi_1)}{2m_1}G\right). \tag{14}$$

Let us now focus on the estimation of f_0, estimation of f_1 is similar. For the father wavelet coefficients, we shall keep the ones defined in Section 4.2. In the expression (14) for f_0 call the first component α_0 and the second β_0; we shall define estimators for their mother wavelet coefficients as

$$\hat{\alpha}_0^{\Psi_{jk}} := \frac{2\hat{\psi}_1^{\Psi_{jk}}}{1+\hat{\phi}_1}\mathbf{1}_{\{\hat{\phi}_1\neq -1\}}, \qquad \hat{\beta}_0^{\Psi_{jk}} := -\left(\frac{1-\hat{\phi}_1}{1+\hat{\phi}_1}\mathbf{1}_{\{\hat{\phi}_1\neq -1\}}\hat{\psi}_1^{\Psi_{jk}} - \frac{\hat{g}(1-\hat{\phi}_1)}{2\hat{m}_1}\mathbf{1}_{\{\hat{m}_1\neq 0\}}\hat{G}^{\Psi_{jk}}\right).$$

Then, what we shall call the 'rough estimator' is defined as:

$$\begin{aligned}\hat{f}_0^R := \sum_{k=0}^{2^{J_n}-1}\hat{f}_0^{\Phi_{J_n k}}\Phi_{J_n k} + \sum_{j=J}^{J_n-1}\sum_{k=0}^{2^j-1}\hat{f}_0^{\Psi_{jk}}\Psi_{jk} \\ + \sum_{j=J_n}^{\tilde{J}_n}\sum_{\ell=0}^{2^j/N-1}\Big(\sum_{k\in\mathfrak{B}_{j\ell}}\hat{\alpha}_0^{\Psi_{jk}}\Psi_{jk}\Big)\mathbf{1}_{\{\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|>\Gamma\sqrt{\log(n)/n}\}} \\ + \sum_{j=J_n}^{\tilde{J}_n}\sum_{\ell=0}^{2^j/N-1}\Big(\sum_{k\in\mathfrak{B}_{j\ell}}\hat{\beta}_0^{\Psi_{jk}}\Psi_{jk}\Big)\mathbf{1}_{\{\|\hat{\beta}_0^{\mathfrak{B}_{j\ell}}\|>\Gamma\hat{T}_n\}},\end{aligned} \tag{15}$$

with $\hat{f}_0^{\Phi_{Jk}}$ and $\hat{f}_0^{\Psi_{jk}}$ as previously defined in Section 4.2 and

$$\hat{T}_n := \sqrt{\frac{\log(n)}{n}}\max\Big(1, \frac{\hat{g}}{|\hat{m}_1|}\mathbf{1}_{\hat{m}_1\neq 0}, \frac{1}{1-\hat{\phi}_1^2}\mathbf{1}_{\hat{\phi}_1^2\neq 1}\Big).$$

It has to be noted that in (15), thresholding of the estimated coefficients of ψ_1 is done "as usual" for density estimation, whereas thresholding of the $\hat{\beta}_0^{\Psi_{jk}}$'s is done with another carefully chosen threshold. The general idea here is that it can be shown that

$$\beta_0 = -\frac{1-\phi_1}{1+\phi_1}f_1, \tag{16}$$

hence if f_1 is much smoother than f_0, we combine the fact that estimation of the stationary density ψ_1 is easy with the other fact that estimating the smoother emission density f_1 leads to a better rate using Theorem 3.

In fact the analysis of the maximum risk of $\hat{f}_0^R$ over the class is much more intricate, and one has to look carefully how the two parts also compensate each other, but at the end we prove that doing so we are able to take advantage of both estimation strenghts. As it was the also the case in Section 3, we also further require a truncation of the estimator to control the risk on events where $\hat{f}_0^R$ may become bad, and we we truncate below at 0 and above at some $\check{T}$, defining

$$\check{f}_0^R := \max\big(0, \min\big(\check{T}, \hat{f}_0^R\big)\big).$$

The following theorem proves that the lower bound (12) is indeed sharp in many interesting regimes.

THEOREM 5. *Suppose* $n\gamma^* \geq \max(\tau^3, \frac{\tau^2 \log(n)^2}{L})$, $\tilde{\jmath}_n > J_n$, $L \leq n$, $\tau \geq 1$, $\check{T} \geq L$, $\zeta \leq 1$, *and* $0 < s_0 \leq S$, *with* $S > 0$ *the regularity of the wavelet basis. Assume* $\tilde{\psi}_2$ *is a given unit vector in* $L^2[0,1]$, *independent of the sample* $Y_1, \dots, Y_n$, *satisfying* $\|\tilde{\psi}_2\|_\infty \leq \tau$ *and* $|\langle \tilde{\psi}_2, \psi_2 \rangle| \geq 7/8$. *Then there are universal constants* $\beta > 0$, $B > 0$ *and* $C > 0$ *such that for all* $\Gamma \geq \beta \max(\frac{L}{\sqrt{\gamma^*}}, \frac{\sqrt{L}}{\tau\gamma^*})$

$$\sup_{\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta \Big(\|\check{f}^R_0 - f_0\|^2_{L^2} \Big) \leq B\check{T}^2 \exp\left(-\frac{Cn\gamma^*\delta^2\epsilon^4\zeta^6}{L^3 + \max(\tau, \sqrt{L})^3 \delta\epsilon^2\zeta^3} \right)$$
$$+ \frac{BL^2}{\delta^2\epsilon^2\zeta^2} \frac{\log(n)}{n\gamma^*} + \frac{BL^3}{\delta^2\epsilon^4\zeta^4} \frac{1}{n\gamma^*} + \frac{B\max(\tau, \sqrt{L})^6}{\delta^2\epsilon^4\zeta^4} \frac{1}{(n\gamma^*)^2} + \frac{R^2}{\min(1, s_0)} \Big(\frac{\Gamma^2}{nR^2\delta^2} \Big)^{2s_0/(2s_0+1)}$$
$$+ \frac{R^2}{\min(1, s_1)} \frac{1}{\delta^2} \Big(\frac{\Gamma^2}{R^2 n\epsilon^2\zeta^2} \Big)^{2s_1/(2s_1+1)} + \frac{BR^2}{\min(1, s_0)} \Big(\frac{\tau^2 \log(n)}{n} \Big)^{2s_0}$$

The proof of Theorem 5 is detailed in Section A.5. As with Theorem 3 and its Corollary 1, of particular interest is the boundary regime, where γ^*, R, L, $\check{T}$ and τ are of constant order while δ, γ and ζ are small, but not too small. The following corollary is intended to illustrate how the bound simplifies in such setting. The proof of Corollary 2 is given in Section A.8.

COROLLARY 2. *Assume that* γ^*, R, L, $\check{T}$, *and* τ *remain constant as* $n \to \infty$ *and* $\delta \geq n^{-a}$, $\epsilon \geq n^{-b}$, $1 \geq \zeta \geq n^{-c}$ *for constants* $a, b, c > 0$ *with* $a, b, c = o(1)$ *as* $n \to \infty$. *Then if* $s_1 < s_0$ *the bound in the Theorem 5 simplifies: for large enough* n,

$$\sup_{\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R) \cap \Sigma_{\gamma^*}(L)} \mathbb{E}_\theta \Big(\|\check{f}^R_0 - f_0\|^2_{L^2} \Big) \leq C \left\{ \frac{1}{\delta^2\epsilon^4\zeta^4 n} + \Big(\frac{1}{\delta^2 n} \Big)^{2s_0/(2s_0+1)} \right\},$$

for a constant C *depending on* γ^*, L, R, Γ, B, τ, $\check{T}$.

In the regime of Corollary 2, *ie*. when δ, ϵ, ζ are small but not too small, if we know that one emission distribution is smoother than the other, and if we are able to get rid of label switching, for example by knowing that the smoothest emission distribution has heavier tails or corresponds to a smaller π_1 (e.g. as in Assumptions 1 or 2 in Proposition 7 of [1]), then Corollary 2 says that we get matching upper and lower bounds. In settings where δ, ϵ, ζ are allowed to be smaller than a polynomial in n, a transition in the rate still occurs according to how s_0 and s_1 compare, but then it may be required to have s_1 much larger than s_0 (depending on δ, ϵ, ζ) to get matching upper and lower bounds.

We believe testing whether f_0 or f_1 is smoother is not possible in general (by comparison to other settings), hence full adaptation is not possible. See also Section 6.

5. Estimation of a separating hyperplane. Theorems 1, 3 and 5 required that $\tilde{\psi}_2$ was a given function independent of the sample, but we need to estimate it. We therefore suppose that $\tilde{\psi}_2$ is estimated using a sample $(\tilde{Y}_1, \dots, \tilde{Y}_n) \sim P^{(n)}_\theta$ that is independent of $(Y_1, \dots, Y_n)$; we discuss in Section 5.2 how to relax this assumption. Similarly to (8), we write $\tilde{\mathbb{P}}^{(s)}_n$ for the empirical distribution of $(\tilde{Y}_1, \dots, \tilde{Y}_n)$.

5.1. *Estimation procedure.* The crude estimator $\tilde{\psi}_2$ is constructed using the heuristic described in Section 2.3. Here again, we use the the S-regular boundary-corrected wavelet basis of [7] with initial resolution level J, see also e.g. [12, Section 4.3.5] and

Section 4. For notational convenience, we define the set of wavelet indices $\Lambda(M) := \{(Jk)_{k=0,\ldots,2^J-1},\ (j,k)_{j=J,\ldots,M,k=0,\ldots,2^j-1}\}$ including all father indices and mother indices up to level $J \le j \le M$, and for all $\lambda \in \Lambda(M)$ we set $e_\lambda = \Phi_{Jk}$ if $\lambda = Jk$ and $e_\lambda = \Psi_{jk}$ if $\lambda = (j,k)$.

For M large enough (see Theorem 6 below) compute the $2^M \times 2^M$ matrix $\tilde{\mathcal{G}}$ with entries

$$\tilde{\mathcal{G}}_{\lambda,\lambda'} = \frac{1}{2}\tilde{\mathbb{P}}_n^{(2)}(e_\lambda \otimes e_{\lambda'} + e_{\lambda'} \otimes e_\lambda) - \tilde{\mathbb{P}}_n^{(1)}(e_\lambda)\tilde{\mathbb{P}}_n^{(1)}(e_{\lambda'}).$$

The matrix $\tilde{\mathcal{G}}$ is an estimator of the matrix $\mathcal{G}$ with entries

$$\mathcal{G}_{\lambda,\lambda'} = \frac{1}{2}\mathbb{E}_\theta(e_\lambda \otimes e_{\lambda'} + e_{\lambda'} \otimes e_\lambda) - \mathbb{E}_\theta(e_\lambda)\mathbb{E}_\theta(e_\lambda) = r(\phi)\langle\psi_2, e_\lambda\rangle\langle\psi_2, e_{\lambda'}\rangle$$

where the second line follows from equation (6). Hence, $\mathcal{G}$ is proportional to the Gram matrix of the vector $V_\theta \propto (\langle\psi_2, e_\lambda\rangle : \lambda \in \Lambda(M))$. The matrices $\tilde{\mathcal{G}}$ and $\mathcal{G}$ are real symmetric, and thus by the spectral theorem are always diagonalizable. By concentration arguments, we expect that $\tilde{\mathcal{G}}$ will have an eigenvalue $\approx r(\phi)$ (which can be positive or negative) and the rest of eigenvalues will be smaller in absolute value. The eigenvector $\tilde{V}$ (chosen such that $\|\tilde{V}\| = 1$) corresponding to the leading eigenvalue is an estimator of $\pm V_\theta/\|V_\theta\|$. We then suggest to use

$$\tilde{\psi}_2(x) := \frac{\max\left(-\tau, \min\left(\tau, \sum_{\lambda\in\Lambda(M)} \tilde{V}_\lambda e_\lambda(x)\right)\right)}{\left(\int_0^1 \max\left(-\tau, \min\left(\tau, \sum_{\lambda\in\Lambda(M)} \tilde{V}_\lambda e_\lambda(y)\right)\right)^2 dy\right)^{1/2}}$$

where the truncation $\tau \ge 1$ is intended to prevent technicalities within the proofs.

THEOREM 6. *Suppose for some $L \ge 1$, $\zeta > 0$, $R > 0$, $s^* > 0$, $M \ge J$ we have*

$$\tau \ge \frac{L}{\zeta}, \qquad 2^{-Ms_*} \le \frac{\zeta\sqrt{2^{2s_*}-1}}{4R}.$$

Then for every $\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)$ with $S \ge \min(s_0, s_1) \ge s_*$,

$$\mathbb{P}_\theta\Big(|\langle\tilde{\psi}_2, \psi_2\rangle| \le \frac{7}{8}\Big) \le 2 \cdot 24^{2^M} \exp\left(-\frac{Cn\gamma^* r(\phi)^2}{L^3 + 2^M\sqrt{L}|r(\phi)|}\right).$$

The proof of Theorem 6 can be found in Section A.6.

5.2. *Discussion about the assumption of two independent samples.* We assumed in the previous that we first get $\tilde{\psi}_2$ based on an independent sample of the HMM. Suppose we are given a single stationary HMM of length $3n$ with distribution $\mathbb{P}_\theta$ such that the hidden Markov chain has absolute spectral gap γ^*. Let $Y' = (Y_1, \ldots, Y_n)$, $\tilde{Y}' = (Y_{2n+1}, \ldots, Y_{3n})$, and denote $\mathbb{P}_{(Y',\tilde{Y}')}$ the distribution of $(Y', \tilde{Y}')$. Denote also $\mathbb{P}_{Y'}$ the distribution of Y' (which is the same as the distribution of $\tilde{Y}'$ by stationarity). For $j = 1, \ldots, 4$ let $\hat{\theta}_j$ denote our estimator of θ_j. Notice that $\hat{\theta}_j$ (resp. θ_j) is non-negative and bounded by 2 (resp. 1) for $j = 1, 2$ and $\check{T}$ (resp. L) for $j = 3, 4$, so that, denoting M (resp. $\tilde{M}$) the upper bound, we have $\|\hat{\theta}_j - \theta_j\| \le M \vee \tilde{M}$, $\|\cdot\|$ being the euclidean norm for $j = 1, 2$ and the $L^2[0,1]$-norm for $j = 3, 4$. Then,

$$\mathbb{E}_{\mathbb{P}_{(Y',\tilde{Y}')}}\left(\|\hat{\theta}_j - \theta_j\|^2\right)$$

$$= \int_0^{M\vee\tilde{M}} \mathbb{P}_{(Y',\tilde{Y}')}\left(\|\hat{\theta}_j - \theta_j\|^2 \ge t\right) dt$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathbb{P}_{Y'}^{\otimes 2}}\left(\|\hat{\theta}_j - \theta_j\|^2\right) + \int_0^{M \vee \tilde{M}} \left[\mathbb{P}_{(Y',\tilde{Y}')}\left(\|\hat{\theta}_j - \theta_j\|^2 \geq t\right) - \mathbb{P}_{Y'}^{\otimes 2}\left(\|\hat{\theta}_j - \theta_j\|^2 \geq t\right)\right] \\
&\leq \mathbb{E}_{\mathbb{P}_{Y'}^{\otimes 2}}\left(\|\hat{\theta}_j - \theta_j\|^2\right) + \left(M \vee \tilde{M}\right)\|\mathbb{P}_{(Y',\tilde{Y}')} - \mathbb{P}_{Y'}^{\otimes 2}\|_{\mathrm{TV}},
\end{aligned}
$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. Using Proposition 1 below, we deduce that the first term on the right side of the last display dominates the second, hence the only cost of using one sample for the whole procedure is a multiplicative constant factor.

PROPOSITION 1. *There exist universal constants C and c such that*

$$
\|\mathbb{P}_{(Y',\tilde{Y}')} - \mathbb{P}_{Y'}^{\otimes 2}\|_{\mathrm{TV}} \leq C e^{-c\gamma^* n}.
$$

PROOF. Denote $Z_i = (X_i, Y_i)$, $i = 1, \dots, 3n$, where $(X_1, \cdots, X_n)$ is the hidden Markov chain. Using similar notations, we have

$$
\|\mathbb{P}_{(Y',\tilde{Y}')} - \mathbb{P}_{Y'}^{\otimes 2}\|_{\mathrm{TV}} \leq \|\mathbb{P}_{(Z',\tilde{Z}')} - \mathbb{P}_{Z'}^{\otimes 2}\|_{\mathrm{TV}}.
$$

Now, for any $(x_1, \dots, x_n, x_{2n+1}, \dots, x_{3n})$, the distribution of $(Y_1, \dots, Y_n, Y_{2n+1}, \dots, Y_{3n})$ conditional on $(X_1, \dots, X_n, X_{2n+1}, \dots, X_{3n}) = (x_1, \dots, x_n, x_{2n+1}, \dots, x_{3n})$ is the same under $\mathbb{P}_{(Y',\tilde{Y}')}$ and $\mathbb{P}_{Y'}^{\otimes 2}$, so that

$$
\|\mathbb{P}_{(Z',\tilde{Z}')} - \mathbb{P}_{Z'}^{\otimes 2}\|_{\mathrm{TV}} \leq 2\|\mathbb{P}_{(X',\tilde{X}')} - \mathbb{P}_{X'}^{\otimes 2}\|_{\mathrm{TV}}
$$

and the result follows from the uniform geometric ergodicity of the binary chain. □

6. Conclusion and open questions. In this paper, we obtain precise behaviour of the minimax risk of all parameters in a nonparametric hidden Markov models, with exact constants regarding the distance to the i.i.d. frontier where the parameters become non-identifiable (we were not interested in the exact dependence of the constants with respect to L, R and γ^*). In particular, we prove a surprising transition in the minimax rates depending on relative smoothnesses of the emission densities.

Similarly to wavelet density estimation with i.i.d. data, the parameter Γ used in the optimal threshold must be chosen depending on the upper L for the supremum norms of f_0, f_1. In the i.i.d. case a simple workaround to adapt to L is to obtain a consistent estimator of the density in L^∞ norm, see [12] Exercise 8.2.1, and plug into the threshold. In the HMM situation, it is not obvious how to obtain an asymptotically valid value for L empirically. Our optimal threshold also depends on γ^*, which requires the preliminary step of the separation hyperplane estimation, itself requiring L. For the estimation of the separating hyperplane, we assume lower bounds on $\min\{s_0, s_1\}$ and on ζ. If neither L nor γ^* is known, the interconnectedness of the parametric and nonparametric part causes us difficulty in fully adapting. We do not know how to build a fully adaptative procedure or if it is even possible.

The main open question concerns full adaptation to get the right constants in the upper bound when a transition occurs relatively to different smoothnesses. We were only able to prove that if the transition exists, then there is an estimator attaining the optimal maximum risk. To be more precise, we did propose pairs of estimators $(\check{f}_0, \check{f}_1)$, $(\check{f}_0^R, \check{f}_1)$, $(\check{f}_0, \check{f}_1^R)$, $(\check{f}_0^R, \check{f}_1^R)$ in which one pair is minimax optimal. When it is known which pair to use, then we indeed get minimax optimal estimators. We tried to build a selection procedure, but we were not able to get good enough upper bounds. We also tried to prove lower bounds to show that full adaptation is not possible but we were not able to. We conjecture that full adaptation is impossible.

Acknowledgments. This work was supported by Institut Universitaire de France, the project ANR-21-CE23-0035-02, and by the EPSRC Programme Grant on the Mathematics of Deep Learning, under the project: EP/V026259/1.

REFERENCES

[1] ABRAHAM, K., CASTILLO, I. and GASSIAT, E. (2022). Multiple testing in nonparametric hidden Markov models: an empirical Bayes approach. *J. Mach. Learn. Res.* **23** Paper No. [94], 57. MR4576679

[2] ABRAHAM, K., GASSIAT, E. and NAULET, Z. (2022). Fundamental limits for learning hidden Markov model parameters. *IEEE Transactions on Information Theory* 1-1. https://doi.org/10.1109/TIT.2022.3213429

[3] ALEXANDROVICH, G., HOLZMANN, H. and LEISTER, A. (2016). Nonparametric identification and maximum likelihood estimation for hidden Markov models. *Biometrika* **103** 423–434.

[4] CAI, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *The Annals of statistics* **27** 898–924.

[5] CAI, T. T. (2008). On information pooling, adaptability and superefficiency in nonparametric function estimation. *Journal of Multivariate Analysis* **99** 421–436.

[6] CHICKEN, E. and CAI, T. T. (2005). Block thresholding for density estimation: local and global adaptivity. *Journal of Multivariate Analysis* **95** 76–106.

[7] COHEN, A., DAUBECHIES, I. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. https://doi.org/10.1006/acha.1993.1005 MR1256527

[8] DE CASTRO, Y., GASSIAT, É. and LACOUR, C. (2016). Minimax adaptive estimation of nonparametric hidden Markov models. *J. Mach. Learn. Res.* **17** Paper No. 111, 43. MR3543517

[9] DE CASTRO, Y., GASSIAT, E. and LE CORFF, S. (2017). Consistent estimation of the filtering and marginal smoothing distributions in nonparametric hidden Markov models. *IEEE Trans. Inform. Theory* **63** 4758–4777.

[10] GASSIAT, É., CLEYNEN, A. and ROBIN, S. (2016). Inference in finite state space non parametric hidden Markov models and applications. *Stat. Comput.* **26** 61–71. MR3439359

[11] GASSIAT, E., ROUSSEAU, J. and VERNET, E. (2018). Efficient semiparametric estimation and model selection for multidimensional mixtures. *Electron. J. Stat.* **12** 703–740. https://doi.org/10.1214/17-EJS1387 MR3769193

[12] GINÉ, E. and NICKL, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. *Cambridge Series in Statistical and Probabilistic Mathematics, [40]*. Cambridge University Press, New York. MR3588285

[13] LE CAM, L. (1986). *Asymptotic methods in statistical decision theory*. Springer-Verlag, New York.

[14] LECESTRE, A. (manuscript). Robust estimation of finite state space hidden Markov models.

[15] LEHÉRICY, L. (2018). State-by-state minimax adaptive estimation for nonparametric hidden Markov models. *J. Mach. Learn. Res.* **19** Paper No. 39, 46. MR3862446

[16] MOSS, D. and ROUSSEAU, J. (2022). Efficient Bayesian estimation and use of cut posterior in semiparametric hidden Markov models.

[17] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** no. 79, 32. MR3383563

[18] TSYBAKOV, A. B. (2009). *Introduction to nonparametric estimation*. Springer-Verlag, New York.

[19] YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963

[20] YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* **102** 315–323. https://doi.org/10.1093/biomet/asv008 MR3371006

APPENDIX A: PROOFS FOR THE UPPER BOUNDS

A.1. Useful lemmas.

LEMMA 1. *The parametrisation $\theta \mapsto (\phi, \psi)$ from (4) is invertible:*

$$\begin{aligned} p &= \tfrac{1}{2}(1-\phi_2)(1-\phi_1), \\ q &= \tfrac{1}{2}(1-\phi_2)(1+\phi_1), \\ f_0 &= \psi_1 - \tfrac{1}{2}\phi_1\phi_3\psi_2 + \tfrac{1}{2}\phi_3\psi_2, \\ f_1 &= \psi_1 - \tfrac{1}{2}\phi_1\phi_3\psi_2 - \tfrac{1}{2}\phi_3\psi_2. \end{aligned}$$

Defining $p_\pm = \frac{1}{2}(1 \mp \tilde{s}\phi_1)(1-\phi_2)$, *where* $\tilde{s} := \mathrm{sgn}(\langle \psi_2, \tilde{\psi}_2\rangle)$ *we have*

$$(p_+, p_-) := \begin{cases} (p,q) & \text{if } \tilde{s} > 0, \\ (q,p) & \text{if } \tilde{s} < 0. \end{cases}$$

Recalling the definition (7) *of* m*, define*

$$g := \phi_3 |\tilde{\mathcal{I}}| = \frac{\sqrt{4m_1^2 m_2 + m_3^2}}{m_2},$$

and define

$$f_\pm := \psi_1 \pm \frac{g(1 \mp \tilde{s}\phi_1)}{2m_1} G, \qquad G := \frac{m_1 \psi_2}{\tilde{\mathcal{I}}}.$$

Then

$$(f_+, f_-) := \begin{cases} (f_0, f_1) & \text{if } \tilde{s} > 0, \\ (f_1, f_0) & \text{if } \tilde{s} < 0. \end{cases}$$

The proof is elementary. Note that $\mathbb{P}_n^{(1)}(\Phi_{Jk})$ is the empirical estimator of $\mathbb{E}_\theta[\Phi_{Jk}] = \langle \Phi_{Jk}, \psi_1\rangle$, hence the above lemma justifies the use of $\hat{f}_0^{\Phi_{Jk}}$, $\hat{f}_1^{\Phi_{Jk}}$ from Section 4.2.

LEMMA 2. *Given* $p^{(3)}_{\phi,\psi}$ *as defined in* (6) *and any function* $\tilde{\psi}$*, one can compute* $m(\phi)$ *defined in* (7)*. Also if* $G = m_1\psi_2/\tilde{\mathcal{I}}$*, then* $\langle \Phi_{Jk}, G\rangle = \mathbb{E}[\tilde{\psi}_2 \otimes \Phi_{Jk}] - \mathbb{E}_\theta[\tilde{\psi}_2]\mathbb{E}_\theta[\Phi_{Jk}]$.

PROOF. We compute, from the expression for $p^{(3)}_{\phi,\psi}$, applied for example to $\tilde{\psi}_2 \otimes 1 \otimes 1$ and using that $\langle \psi_1, 1\rangle = \int \psi_1 = 1$, $\langle \psi_2, 1\rangle = 0$,

$$\begin{aligned}
\mathbb{E}_\theta(\tilde{\psi}_2) &= \langle \psi_1, \tilde{\psi}_2\rangle \\
\mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2) &= \langle \psi_1, \tilde{\psi}_2\rangle^2 + r(\phi)\langle \psi_2, \tilde{\psi}_2\rangle^2 \\
\mathbb{E}_\theta(\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2) &= \langle \psi_1, \tilde{\psi}_2\rangle^2 + r(\phi)\phi_2\langle \psi_2, \tilde{\psi}_2\rangle^2 \\
\mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2) &= \langle \psi_1, \tilde{\psi}_2\rangle^3 + (2r(\phi) + r(\phi)\phi_2)\langle \psi_2, \tilde{\psi}_2\rangle^2\langle \psi_1, \tilde{\psi}_2\rangle - r(\phi)\phi_1\phi_2\phi_3\langle \psi_2, \tilde{\psi}_2\rangle^3
\end{aligned}$$

Then $m := (r(\phi)\tilde{\mathcal{I}}^2, r(\phi)\phi_2\tilde{\mathcal{I}}^2, r(\phi)\phi_1\phi_2\phi_3\tilde{\mathcal{I}}^3)$, $\tilde{\mathcal{I}} := \langle \psi_2, \tilde{\psi}_2\rangle$ is easily extracted.

Similarly, $\mathbb{E}_\theta[\tilde{\psi}_2 \otimes \Phi_{Jk}] = \langle \psi_1, \tilde{\psi}_2\rangle\langle \psi_1, \Phi_{Jk}\rangle + r(\phi)\tilde{\mathcal{I}}\langle \psi_2, \Phi_{Jk}\rangle$, and the expression for the coefficient of G can be extracted. □

The following bounds are immediate from the definition of the parameter space (3) and the reparametrisation (4) (recall also the definition (5) of r).

LEMMA 3. *For* ϕ *corresponding to* $\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ *we have the bounds*

$$-\frac{1-\delta}{1+\delta} \le \phi_1 \le \frac{1-\delta}{1+\delta}, \quad \epsilon \le |\phi_2| \le 1-2\delta, \quad \phi_3 \ge \zeta, \quad \delta\epsilon\zeta^2/4 \le |r(\phi)| \le \phi_3^2/4.$$

LEMMA 4. *Let* m_1, m_2, m_3 *be defined as in* (7) *and let* $v := 4m_1^2 m_2 + m_3^2$*. Then* $0 \le m_2 \le |m_1|$ *and* $\sqrt{v} = \tilde{\mathcal{I}}^3 r(\phi)\phi_2\phi_3 = \tilde{\mathcal{I}} m_2\phi_3$*. Furthermore, for every* $\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ *and* $0 < \delta \le 1$, $0 < \epsilon \le 1$*, and* $0 < \zeta \le 1$*:*

$$\left|\frac{g}{m_1}\right| \le \frac{4}{\delta\epsilon\zeta|\tilde{\mathcal{I}}|}, \qquad \frac{\max(1,g)}{m_2} \le \frac{4}{\delta\epsilon^2\zeta^2|\tilde{\mathcal{I}}|^2}, \qquad \frac{\max(1,g)}{gm_2} \le \frac{4}{\delta\epsilon^2\zeta^3|\tilde{\mathcal{I}}|^3}.$$

PROOF. Observe that $m_2 = m_1\phi_2$ and $|\phi_2| \leq 1$. Also, $m_2 = r(\phi)\phi_2\tilde{\mathcal{I}}^2 = \frac{1}{4}(1-\phi_1^2)\phi_2^2\phi_3^2\tilde{\mathcal{I}}^2 \geq 0$. Similarly,

$$v = 4r(\phi)^2\tilde{\mathcal{I}}^4 \cdot r(\phi)\phi_2\tilde{\mathcal{I}}^2 + r(\phi)^2\phi_1^2\phi_2^2\phi_3^2\tilde{\mathcal{I}}^6 = r(\phi)^2\tilde{\mathcal{I}}^6\Big(4r(\phi)\phi_2 + \phi_1^2\phi_2^2\phi_3^2\Big) = r(\phi)^2\phi_2^2\phi_3^2\tilde{\mathcal{I}}^6.$$

Next, observe that $\frac{g}{m_1} = \frac{\phi_3|\tilde{\mathcal{I}}|}{\frac{1}{4}(1-\phi_1^2)\phi_2\phi_3^2|\tilde{\mathcal{I}}|^2} = \frac{4}{(1-\phi_1^2)\phi_2\phi_3|\tilde{\mathcal{I}}|}$. But $0 \geq 1-\phi_1^2 \geq \frac{4\delta}{(1+\delta)^2} \geq \delta$, $|\phi_2| \geq \epsilon$, and $\phi_3 \geq \zeta$ by Lemma 3. Similarly, since $g = \phi_3|\tilde{\mathcal{I}}| \leq \zeta \leq 1$, $0 \leq \frac{\max(1,g)}{m_2} = \frac{1}{m_2} = \frac{4}{(1-\phi_1^2)\phi_2^2\phi_3^2|\tilde{\mathcal{I}}|^2} \leq \frac{4}{\delta\epsilon^2\zeta^2|\tilde{\mathcal{I}}|^2}$. □

LEMMA 5. *For any $k \geq 1$,*

$$\|p_\theta^{(k)}\|_\infty \leq \max\big(\|f_0\|_\infty,\ \|f_1\|_\infty\big)^k.$$

Consequently, for any $\theta \in \Sigma_{\gamma^*}(L)$ and any measurable function $h : \mathbb{R}^k \to \mathbb{R}$, we have

$$\mathbb{E}_\theta[h(Y_1,\ldots,Y_k)^2] \leq L^k\|h\|_{L^2}^2.$$

PROOF. Observe that $p_\theta^{(k)}(y_1,\ldots,y_k) = \sum_{x_1,\ldots,x_k}\mathbb{P}_\theta(X_1 = x_1,\ldots,X_k = x_k)\prod_{i=1}^k f_{x_i}(y_i)$. The first conclusion is immediate, and the second follows from

$$\mathbb{E}_\theta h(Y_1,\ldots,Y_k)^2 = \int p_\theta^{(k)}(y_1,\ldots,y_k)h(y_1,\ldots,y_k)\mathrm{d}y_1\cdots\mathrm{d}y_k \leq \|p^{(k)}\|_\infty\|h\|_{L^2}^2. \quad \square$$

REMARK 1. *The proof adapts to yield $E_\theta[h(Y_1,Y_3)^2] \leq L^2\|h\|_{L^2}^2$ rather than the weaker bound $L^3\|h\|_{L^2}^2$ directly obtainable using the lemma. Indeed, we have*

$$\sup_{y_1,y_3}\Big|\int p^{(3)}(y_1,y_2,y_3)\mathrm{d}y_2\Big| = \sum_{x_1,x_2,x_3}\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, X_3 = x_3)f_{x_1}(y_1)f_{x_3}(y_3) \leq L^2,$$

and the rest of the proof is the same.

LEMMA 6. *For all $\theta \in \Sigma_{\gamma^*}(L)$, $\phi_3 \leq \sqrt{2L}$.*

PROOF. We compute $\phi_3^2 = \int_0^1(f_0-f_1)^2 \leq \|f_0-f_1\|_\infty\int_0^1(|f_0|+|f_1|) = 2\|f_0-f_1\|_\infty$. Since we have the pointwise bounds $0 \leq f_0, f_1 \leq L$ for every $\theta \in \Sigma_{\gamma^*}(L)$, it follows that $\phi_3^2 \leq 2L$. We remark that this upper bound is tight since it is attained for instance when f_0 is the uniform density on $[0,1/L]$ and f_1 the uniform density on $[1-1/L,1]$. □

We now recall the following result, which is adapted from [17] and will be key to getting deviation inequalities of empirical ingredients in our procedures.

LEMMA 7. *Let $1 \leq k \leq 3$ and let $h : \mathbb{R}^k \to \mathbb{R}$ be measurable. There is a universal constant $C > 0$ such that for all θ, all $n \geq 4$ such that $n\gamma^* \geq 1/99$, and all $t \geq 0$*

$$\mathbb{P}_\theta\Big(|\mathbb{P}_n^{(k)}(h) - \mathbb{E}_\theta(h)| \geq t\Big) \leq \exp\Big(-\frac{Cnt^2\gamma^*}{\mathbb{E}_\theta(h^2) + \|h\|_\infty t}\Big).$$

This in particular implies that there is a is a universal constant $C > 0$ such that for all θ, all $n \geq 4$ such that $n\gamma^* \geq 1/99$, and all $x \geq 0$

$$\mathbb{P}_\theta\left(|\mathbb{P}_n^{(k)}(h) - \mathbb{E}_\theta(h)| \geq C\sqrt{\frac{\mathbb{E}_\theta[h^2]x}{n\gamma^*}} + \frac{C\|h\|_\infty x}{n\gamma^*}\right) \leq e^{-x}.$$

PROOF. Since $1 \leq k \leq 3$, we can view any function $h : \mathbb{R}^k \to \mathbb{R}$ as $\tilde{h} : \mathbb{R}^6 \to \mathbb{R}$ with $h(Y_i, \ldots, Y_{i+k}) = \tilde{h}(X_i, X_{i+1}, X_{i+2}, Y_i, Y_{i+1}, Y_{i+2})$. By our assumptions, the process $\big((X_i, X_{i+1}, X_{i+2}, Y_i, Y_{i+1}, Y_{i+2})\big)_{i\geq 1}$ is a stationary Markov Chain with pseudo spectral gap (defined as in [17]) $\gamma_{\mathrm{ps}} \geq \gamma^*/8$. Indeed, calculations in [2, Lemma 1] based on the relationship between the pseudo spectral gap and the mixing time show that $\gamma_{\mathrm{ps}} \geq 0.5((\log 4/\gamma^*) + 2)^{-1}$, and the bound $\max(\gamma^*, \log 2) \leq 1$ yields the claimed bound.

By Theorem 3.4 in [17] (though note there is an updated version of the paper on arXiv), for $S_n := \sum_{i=1}^{n-k+1} \tilde{h}(X_i, X_{i+1}, X_{i+2}, Y_i, Y_{i+1}, Y_{i+2})$ we do have for any $t \geq 0$

$$\mathbb{P}_\theta\Big(|S_n - \mathbb{E}_\theta(S_n)| \geq t\Big) \leq \exp\Big(-\frac{t^2\gamma_{\mathrm{ps}}}{8(n-k+1+1/\gamma_{\mathrm{ps}})\mathbb{E}_\theta(h^2) + 20\|h\|_\infty t}\Big).$$

Dividing S_n by $n-k+1$ and replacing $n-k+1$ and γ_{ps} by the respective lower bounds $n/2$ and $\gamma^*/8$, we find that

$$\begin{aligned}\mathbb{P}_\theta\Big(|\mathbb{P}_n^{(k)}(h) - \mathbb{E}_\theta(h)| \geq t\Big) &\leq \exp\Big(-\frac{nt^2\gamma^*/16}{8(1+\frac{16}{n\gamma^*})\mathbb{E}_\theta(h^2) + 20\|h\|_\infty t}\Big) \\ &\leq \exp\Big(-\frac{nt^2\gamma^*}{16\times 8\times(1+16\times 99)\times\mathbb{E}_\theta(h^2) + 320\|h\|_\infty t}\Big)\end{aligned}$$

under the assumption that $n\gamma^* \geq 1/99$. The result follows by taking $t = C\sqrt{\mathbb{E}_\theta[h^2]x/(n\gamma^*)} + C\|h\|_\infty x/(n\gamma^*)$ for C a sufficiently large constant that the argument of the exponential is smaller than $-x$ (by splitting into cases based on which of the two terms in the denominator is larger it can be seen that it suffices to take $C = \max(\sqrt{2\times 16\times 8\times(1+16\times 99)}, 640) = 640$), yielding the claim. $\square$

The following consequence of deviation inequalities to get bounds in expectation will also be used.

LEMMA 8. *Suppose X is a non-negative random variable and there exist $a, b, c > 0$ such that $\mathbb{P}(X > b\sqrt{x/n} + ax/n) \leq ce^{-x}$ for all $x > 0$. There for all $d \geq 0$*

$$\mathbb{E}(X^2\mathbf{1}_{\{X>d\}}) \leq c\Big(d^2 + \frac{5b^2}{4n} + \frac{7a^2}{2n^2}\Big)\exp\Big(-\frac{nd^2}{2b^2+8ad}\Big).$$

PROOF. Applying the standard identity $\mathbb{E}(Y) = \int_0^\infty \mathbb{P}(Y > y)\mathrm{d}y$ for any non-negative random variable Y to $Y = X^2\mathbf{1}_{\{X>d\}}$ and making the substitution $y = u^2$ we obtain

$$\begin{aligned}\mathbb{E}(X^2\mathbf{1}_{\{X>d\}}) &= \int_0^\infty \mathbb{P}\big(X^2\mathbf{1}_{\{X>d\}} > y\big)\mathrm{d}y \\ &= \int_0^\infty \mathbb{P}\big(X > \max(d, \sqrt{y})\big)\mathrm{d}y \\ &= \int_0^{d^2} \mathbb{P}(X > d)\mathrm{d}y + \int_{d^2}^\infty \mathbb{P}\big(X > \sqrt{y}\big)\mathrm{d}y \\ &= d^2\mathbb{P}(X > d) + \int_d^\infty 2u\mathbb{P}(X > u)\mathrm{d}u.\end{aligned}$$

Define $\varphi(x) := \frac{b}{2a}\big(\sqrt{1+4ax/b^2} - 1\big)$. For the change of variables $u = b\sqrt{x/n} + ax/n$ one calculates that $x = n\varphi(u)^2$ and hence computes, using Cauchy–Schwarz for the penultimate line,

$$\begin{aligned}\int_d^\infty u\mathbb{P}(X>u)\mathrm{d}u &= \int_{n\varphi(d)^2}^\infty \Big(b\sqrt{\frac{x}{n}}+a\frac{x}{n}\Big)\Big(\frac{b}{2\sqrt{nx}}+\frac{a}{n}\Big)\mathbb{P}\Big(X>b\sqrt{\frac{x}{n}}+a\frac{x}{n}\Big)\mathrm{d}x\\
&\le c\int_{n\varphi(d)^2}^\infty \Big(\frac{b^2}{2n}+\frac{3}{2}\frac{b}{\sqrt{n}}\frac{a\sqrt{x}}{n}+\frac{a^2x}{n^2}\Big)e^{-x}\mathrm{d}x\\
&\le c\int_{n\varphi(d)^2}^\infty \Big(\frac{5b^2}{4n}+\frac{7a^2x}{4n^2}\Big)e^{-x}\mathrm{d}x\\
&= \frac{c}{4}\Big(\frac{5b^2}{n}+\frac{7a^2}{n^2}(n\varphi(d)^2+1)\Big)e^{-n\varphi(d)^2}.\end{aligned}$$

Similarly one has

$$\mathbb{P}(X>d)=\mathbb{P}\Big(X>b\sqrt{\frac{n\varphi(d)^2}{n}}+a\frac{n\varphi(d)^2}{n}\Big)\le ce^{-n\varphi(d)^2}.$$

To obtain the final expression, we remark that $xe^{-x}\le \frac{2}{e}e^{-x/2}$, that $2/e+1\le 2$ and that for all $x>0$

$$\varphi(x)\ge\frac{b}{2a}\frac{4ax/b^2}{2\sqrt{1+4ax/b^2}}=\frac{x/b}{\sqrt{1+4ax/b^2}}. \qquad \square$$

A.2. Solving the direct problem: inequalities for the m functional. Recall the definitions

$$\begin{aligned}\hat m_1 &:= \mathbb{P}_n^{(2)}(\tilde\psi_2\otimes\tilde\psi_2)-\mathbb{P}_n^{(1)}(\tilde\psi_2)^2,\\
\hat m_2 &:= \mathbb{P}_n^{(3)}(\tilde\psi_2\otimes 1\otimes\tilde\psi_2)-\mathbb{P}_n^{(1)}(\tilde\psi_2)^2\\
\hat m_3 &= -\mathbb{P}_n^{(3)}(\tilde\psi_2\otimes\tilde\psi_2\otimes\tilde\psi_2)+\mathbb{P}_n^{(1)}(\tilde\psi_2)^3+\big(2\hat m_1+\hat m_2\big)\mathbb{P}_n^{(1)}(\tilde\psi_2),\end{aligned}$$

estimators of the functional m defined in (7) as $m=(r(\phi)\tilde{\mathcal{I}}^2, r(\phi)\phi_2\tilde{\mathcal{I}}^2, r(\phi)\phi_1\phi_2\phi_3\tilde{\mathcal{I}}^3)$ with $\tilde{\mathcal{I}}=\langle\psi_2,\tilde\psi_2\rangle$, and deduced from Lemma 2 to be equal to what is obtained in the expressions for $\hat m$ on replacing every instance of an empirical estimator by the expectation operator. [This does not mean that $\mathbb{E}_\theta\hat m=m$, since there are powers and products in the expressions.] In this section, we prove deviation inequalities for the estimators of m, from which we deduce bounds in expectation. The results of this section will be used to prove Theorem 1 and Theorem 3.

We remark that the results are mostly uniform over the whole class $\Sigma_{\gamma^*}(L)$, not our final parameter set $\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)\cap\Sigma_{\gamma^*}(L)$. The need to intersect with $\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$ arises for ensuring the parameters θ are identifiable from m.

PROPOSITION 2. *Let $n\gamma^*\ge 1/99$. Then there exists a universal constant $C>0$ such that for all $x\ge 0$*

$$\sup_{\theta\in\Sigma_{\gamma^*}(L)}\mathbb{P}_\theta\left(\max_{j=1,2}|\hat m_j-m_j|\ge CL\sqrt{\frac{x}{n\gamma^*}}+C\max(\tau,\sqrt{L})^2\frac{x}{n\gamma^*}\right)\le 3e^{-x}.$$

PROPOSITION 3. *Let $n\gamma^*\ge 1/99$. Then there exists a universal constant $C>0$ such that for all $x\ge 0$*

$$\sup_{\theta\in\Sigma_{\gamma^*}(L)}\mathbb{P}_\theta\left(\max_{j=1,2,3}|\hat m_j-m_j|\ge CL^{3/2}\sqrt{\frac{x}{n\gamma^*}}+C\max(\tau,\sqrt{L})^3\frac{x}{n\gamma^*}\right)\le 4e^{-x}.$$

PROPOSITION 4. *There exists a constant $K>0$ such that whenever $n\gamma^*\geq 1/99$,*

$$\sup_{\theta\in\Sigma_{\gamma^*}(L)} \mathbb{E}_\theta\Big(\max_{j=1,2,3}|\hat{m}_j-m_j|^2\Big)\leq K\Big(\frac{L^3}{n\gamma^*}+\frac{\max(\tau,\sqrt{L})^6}{(n\gamma^*)^2}\Big).$$

PROPOSITION 5. *Assume $n\gamma^*\geq 1/99$, $|\tilde{\mathcal{I}}|\geq 7/8$ and $\zeta\leq 1$, and define the event*

$$\Omega_n := \Big\{\max_{j=1,2}\Big|\frac{\hat{m}_j}{m_j}-1\Big|\leq\frac{1}{2},\ \max_{j=1,2,3}|\hat{m}_j-m_j|\leq\frac{gm_2}{44\max(1,g)}\Big\}. \tag{17}$$

Then there exists a universal constant $C>0$ such that

$$\sup_{\theta\in\Sigma_{\gamma^*}(L)}\mathbb{P}_\theta(\Omega_n^c)\leq 7\exp\left(-\frac{Cn\gamma^* g^2m_2^2/\max(1,g)^2}{L^3+\max(\tau,\sqrt{L})^3gm_2/\max(1,g)}\right),$$

$$\sup_{\theta\in\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)\cap\Sigma_{\gamma^*}(L)}\mathbb{P}_\theta(\Omega_n^c)\leq 7\exp\left(-\frac{Cn\gamma^*\delta^2\epsilon^4\zeta^6}{L^3+\max(\tau,\sqrt{L})^3\delta\epsilon^2\zeta^3}\right).$$

The proof of Proposition 3 is the most involved of these, and we outline how to prove the other results before addressing it.

PROOF OF PROPOSITION 2. The proof is similar to the proof of Proposition 3, where $\max_{j=1,2,3}|\hat{m}_j-m_j|$ is controlled. Here, since only $\hat{m}_1$ and $\hat{m}_2$ are involved, the proxy variance is no more than L since only $\mathbb{P}_n^{(2)}$ is involved (versus $L^{3/2}$ when $\mathbb{P}_n^{(3)}$ is involved). $\square$

PROOF OF PROPOSITION 4. In view of Proposition 3 we may apply Lemma 8 with $a=C\max(\tau,\sqrt{L})^3/\gamma^*$, $b=CL^{3/2}/\sqrt{\gamma^*}$, $c=8$ and $d=0$ to obtain the claimed bound. $\square$

PROOF OF PROPOSITION 5. The first inequality essentially follows from Propositions 2 and 3 and a change of variables: see Lemmas 10 and 11 (and the sentence after the former) below where this change of variables is explicitly made. The second inequality follows from the fact that $\frac{\max(1,g)}{gm_2}\leq\frac{16}{\delta\epsilon^2\zeta^3\tilde{\mathcal{I}}^2}$ on $\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$ by Lemma 4. $\square$

PROOF OF PROPOSITION 3. We have that $\max_{j=1,2,3}|\hat{m}_j-m_j|\leq 16\|\tilde{\psi}_2\|_\infty^3\leq 16\tau^3$ by construction. Hence whenever $x>n\gamma^*$ we have with probability $1\geq 1-e^{-x}$ under $\mathbb{P}_\theta$ that

$$\max_{j=1,2,3}|\hat{m}_j-m_j|\leq 16\tau^3\leq CL^{3/2}\sqrt{\frac{x}{n\gamma^*}}+C\max(\tau,\sqrt{L})^3\frac{x}{n\gamma^*}$$

Next we address the case $x\leq n\gamma^*$. It is seen that

$$\hat{m}_1-m_1=\mathbb{P}_n^{(2)}(\tilde{\psi}_2\otimes\tilde{\psi}_2)-\mathbb{E}_\theta(\tilde{\psi}_2\otimes\tilde{\psi}_2)-\Big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2)^2-\mathbb{E}_\theta(\tilde{\psi}_2)^2\Big)$$

ie.

$$\hat{m}_1-m_1=\Big(\mathbb{P}_n^{(2)}(\tilde{\psi}_2\otimes\tilde{\psi}_2)-\mathbb{E}_\theta(\tilde{\psi}_2\otimes\tilde{\psi}_2)\Big)-2\mathbb{E}_\theta(\tilde{\psi}_2)\Big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2)-\mathbb{E}_\theta(\tilde{\psi}_2)\Big)-\Big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2)-\mathbb{E}_\theta(\tilde{\psi}_2)\Big)^2.$$

Noting that $\mathbb{E}_\theta(|\tilde{\psi}_2|)\leq\mathbb{E}_\theta(\tilde{\psi}_2^2)^{1/2}\leq\sqrt{L}\|\tilde{\psi}_2\|_{L^2}=\sqrt{L}$ whenever $\theta\in\Sigma_{\gamma^*}(L)$ by Lemma 5, we deduce

$$|\hat{m}_1-m_1|\leq|Z_2|+2\sqrt{L}|Z_1|+|Z_1|^2,$$

where $Z_1 = \mathbb{P}_n^{(1)}(\tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2)$ and $Z_2 = \mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2)$. The same reasoning yields, with , $Z_3 = \mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2 \otimes 1 \otimes \tilde{\psi}_2)$,

$$|\hat{m}_2 - m_2| \leq |Z_3| + 2\sqrt{L}|Z_1| + |Z_1|^2.$$

The decomposition for $\hat{m}_3 - m_3$ is similar but slightly more involved. Since $m_3 = -\mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2) + \mathbb{E}_\theta(\tilde{\psi}_2)^3 + \big(2m_1 + m_2\big)\mathbb{E}_\theta(\tilde{\psi}_2)$, we deduce

$$\begin{aligned}
\hat{m}_3 - m_3 = &-\Big(\mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2)\Big) \\
&+ \mathbb{P}_n^{(1)}(\tilde{\psi}_2)^3 - \mathbb{E}_\theta(\tilde{\psi}_2)^3 \\
&+ \big[(2\hat{m}_1 + \hat{m}_2) - (2m_1 + m_2)\big]\mathbb{E}_\theta(\tilde{\psi}_2) \\
&+ (2m_1 + m_2)\big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2)\big) \\
&+ \big[(2\hat{m}_1 + \hat{m}_2) - (2m_1 + m_2)\big]\big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2)\big).
\end{aligned}$$

But $\mathbb{P}_n^{(1)}(\tilde{\psi}_2)^3 - \mathbb{E}_\theta(\tilde{\psi}_2)^3 = 3\mathbb{E}_\theta(\tilde{\psi}_2)^2 Z_1 + 3\mathbb{E}_\theta(\tilde{\psi}_2) Z_1^2 + Z_1^3$, and thus recalling $\mathbb{E}_\theta(|\tilde{\psi}_2|) \leq \sqrt{L}$ and $m_2 \leq |m_1| \leq \frac{1}{4}\phi_3^2 \leq \frac{1}{2}L$ by Lemmas 6 and 3, writing $Z_4 = \mathbb{P}_n^{(3)}(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \tilde{\psi}_2 \otimes \tilde{\psi}_2)$ we have

$$\begin{aligned}
|\hat{m}_3 - m_3| \leq &|Z_4| + 3L|Z_1| + 3\sqrt{L}|Z_1|^2 + |Z_1|^3 + 2\sqrt{L}|\hat{m}_1 - m_1| + \sqrt{L}|\hat{m}_2 - m_2| \\
&+ \frac{3L}{2}|Z_1| + 2|\hat{m}_1 - m_1||Z_1| + |\hat{m}_2 - m_2||Z_1|.
\end{aligned}$$

It follows (recall $L \geq 1$ necessarily)

$$\begin{aligned}
\max_{j=1,2,3} |\hat{m}_j - m_j| \leq &|Z_4| + \sqrt{L}|Z_3| + 2\sqrt{L}|Z_2| + 10.5L|Z_1| \\
&+ 9\sqrt{L}Z_1^2 + 4|Z_1|^3 + 2|Z_1 Z_2| + |Z_1 Z_3|.
\end{aligned}$$

Feeding in bounds on the Z_i from Lemma 9 below, we deduce with probability at least $1 - 4e^{-x}$ under $\mathbb{P}_\theta$ that

$$\begin{aligned}
\max_{j=1,2,3} |\hat{m}_j - m_j| \leq &C\left(L^{3/2}\sqrt{\frac{x}{n\gamma^*}} + \tau^3 \frac{x}{n\gamma^*}\right) + 3C\left(L^{3/2}\sqrt{\frac{x}{n\gamma^*}} + L^{1/2}\tau^2 \frac{x}{n\gamma^*}\right) \\
&+ 10.5C\left(L^{3/2}\sqrt{\frac{x}{n\gamma^*}} + L\tau \frac{x}{n\gamma^*}\right) + 9C^2\sqrt{L}\left(L^{1/2}\sqrt{\frac{x}{n\gamma^*}} + \tau \frac{x}{n\gamma^*}\right)^2 \\
&+ 4C^3\left(L^{1/2}\sqrt{\frac{x}{n\gamma^*}} + \tau \frac{x}{n\gamma^*}\right)^3 \\
&+ 3C^2\left(L^{1/2}\sqrt{\frac{x}{n\gamma^*}} + \tau \frac{x}{n\gamma^*}\right)\left(L\sqrt{\frac{x}{n\gamma^*}} + \tau^2 \frac{x}{n\gamma^*}\right).
\end{aligned}$$

Grouping together the terms with same powers, still with probability at least $1 - 8e^{-x}$ under $\mathbb{P}_\theta$

$$\max_{j=1,2,3} |\hat{m}_j - m_j| \leq 14.5CL^{3/2}\Big(\frac{x}{n\gamma^*}\Big)^{1/2} + C\Big(\tau^3 + 3L^{1/2}\tau^2 + 10.5L\tau + 12CL^{3/2}\Big)\frac{x}{n\gamma^*}$$

$$+ C^2\Big(18\tau L + 4CL^{3/2} + 3\tau^2\sqrt{L} + 3\tau L\Big)\Big(\frac{x}{n\gamma^*}\Big)^{3/2}$$
$$+ C^2\Big(9\sqrt{L}\tau^2 + 12C\tau L + 3\tau^3\Big)\Big(\frac{x}{n\gamma^*}\Big)^2 + 12C^3\tau^2\sqrt{L}\Big(\frac{x}{n\gamma^*}\Big)^{5/2}$$
$$+ 4C^3\tau^3\Big(\frac{x}{n\gamma^*}\Big)^3.$$

The conclusion follows since we are in the case where $x \le n\gamma^*$, and because $L \ge 1$ and $\tau \ge 1$. $\square$

LEMMA 9. *Assume $\theta \in \Sigma_{\gamma^*}(L)$ and $n\gamma^* \ge 1/99$. Write $Z_1 = \mathbb{P}_n^{(1)}(\tilde\psi_2) - \mathbb{E}_\theta(\tilde\psi_2)$, $Z_2 = \mathbb{P}_n^{(2)}(\tilde\psi_2 \otimes \tilde\psi_2) - \mathbb{E}_\theta(\tilde\psi_2 \otimes \tilde\psi_2)$, $Z_3 = \mathbb{P}_n^{(3)}(\tilde\psi_2 \otimes 1 \otimes \tilde\psi_2) - \mathbb{E}_\theta(\tilde\psi_2 \otimes 1 \otimes \tilde\psi_2)$, and $Z_4 = \mathbb{P}_n^{(3)}(\tilde\psi_2 \otimes \tilde\psi_2 \otimes \tilde\psi_2) - \mathbb{E}_\theta(\tilde\psi_2 \otimes \tilde\psi_2 \otimes \tilde\psi_2)$. Then*

$$\mathbb{P}_\theta\Big(|Z_1| \ge C\sqrt{\frac{Lx}{n\gamma^*}} + C\tau\frac{x}{n\gamma^*}\Big) \le e^{-x},$$
$$\mathbb{P}_\theta\Big(|Z_j| \ge CL\sqrt{\frac{x}{n\gamma^*}} + C\tau^2\frac{x}{n\gamma^*}\Big) \le e^{-x}, \quad j = 2, 3,$$
$$\mathbb{P}_\theta\Big(|Z_4| \ge CL^{3/2}\sqrt{\frac{x}{n\gamma^*}} + C\tau^3\frac{x}{n\gamma^*}\Big) \le e^{-x}.$$

PROOF. For Z_4, use Lemma 7 together with the facts that $\|\tilde\psi_2 \otimes \tilde\psi_2 \otimes \tilde\psi_2\|_\infty = \|\tilde\psi_2\|_\infty^3 \le \tau^3$ and that $\mathbb{E}_\theta[(\tilde\psi_2 \otimes \tilde\psi_2 \otimes \tilde\psi_2)^2] \le L^3\|\tilde\psi_2\|_{L^2}^6 = L^3$ by Lemma 5. The arguments are similar for $j = 1, 2, 3$, though note for $j = 3$ we use Remark 1 rather than Lemma 5 itself. $\square$

LEMMA 10. *Let $n\gamma^* \ge 1/99$. Then, there exists a universal constant $C > 0$ such that for all $\theta \in \Sigma_{\gamma^*}(L)$*

$$\mathbb{P}_\theta\Big(\max_{j=1,2}\Big|\frac{\hat m_j}{m_j} - 1\Big| \ge \frac{1}{2}\Big) \le 3\exp\left(-\frac{Cn\gamma^* m_2^2}{L^2 + \max(\tau, \sqrt{L})^2 m_2}\right).$$

Note that $\frac{gm_2}{\max(1,g)} \le m_2$ and that $L \ge 1$ necessarily, hence the the absolute value of the exponent in Lemma 10 is larger than that in Lemma 11.

PROOF. We apply Proposition 2 with $x \ge 0$ such that

$$CL\sqrt{\frac{x}{n\gamma^*}} + C\max(\tau, \sqrt{L})^2\frac{x}{n\gamma^*} = \frac{m_2}{2},$$

i.e.,

$$\sqrt{\frac{x}{n\gamma^*}} = \frac{L}{2\max(\tau, \sqrt{L})^2}\Big(\sqrt{1 + \frac{2\max(\tau, \sqrt{L})^2 m_2}{CL^2}} - 1\Big)$$
$$\ge \frac{L}{2}\frac{m_2/(CL^2)}{\sqrt{1 + \frac{2\max(\tau,\sqrt{L})^2 m_2}{CL^2}}}.$$

Then, using that $0 \le m_2 \le |m_1|$, (Lemma 4), we have

$$\mathbb{P}_\theta\Big(\max_{j=1,2}\Big|\frac{\hat m_j}{m_j}-1\Big| \ge \frac{1}{2}\Big) \le \mathbb{P}_\theta\Big(\max_{j=1,2}\Big|\hat m_j - m_j\Big| \ge \frac{m_2}{2}\Big)$$
$$\le 6\exp\left(-\frac{n\gamma^* m_2^2}{2C^2L^2 + 2C\max(\tau,\sqrt{L})^2|m_2|}\right)$$

concluding the proof. ☐

LEMMA 11. *Let $n\gamma^* \ge 1/99$. Then, there exists a universal constant $C>0$ such that for all $\theta \in \Sigma_{\gamma^*}(L)$*

$$\mathbb{P}_\theta\Big(\max_{j=1,2,3}|\hat m_j - m_j| \ge \frac{gm_2}{44\max(1,g)}\Big) \le 4\exp\left(-\frac{Cn\gamma^* g^2 m_2^2/\max(1,g)^2}{L^3 + \max(\tau,\sqrt{L})^3 gm_2/\max(1,g)}\right).$$

PROOF. By Proposition 3, applied with $x \ge 0$ such that

$$CL^{3/2}\sqrt{\frac{x}{n\gamma^*}} + C\max(\tau,\sqrt{L})^3\frac{x}{n\gamma^*} = \frac{gm_2}{44\max(1,g)}$$

ie,

$$\sqrt{\frac{x}{n\gamma^*}} = \frac{L^{3/2}}{2\max(\tau,\sqrt{L})^3}\left(\sqrt{1+\frac{4\max(\tau,\sqrt{L})^3 gm_2}{44CL^3\max(1,g)}}-1\right)$$
$$\ge \frac{1}{44CL^{3/2}}\frac{gm_2/\max(1,g)}{\sqrt{1+\frac{4\max(\tau,\sqrt{L})^3 gm_2}{44CL^3\max(1,g)}}},$$

we obtain the result. ☐

A.3. Proof of Theorem 1. Due to label switching, $\hat\phi_1$ may be either an estimator of ϕ_1 or $-\phi_1$, depending on the value of $\tilde s := \mathrm{sgn}(\langle\psi_2,\tilde\psi_2\rangle)$. In the proofs, rather than allow an arbitrary permutation, we define $p_\pm$ as an (unobserved) permutation of (p,q) and we define $\hat p_+, \hat p_-$ such that $\hat p_\pm$ estimates $p_\pm$. To this end, define $p_\pm = \frac{1}{2}(1\mp\tilde s\phi_1)(1-\phi_2)$ (as in Lemma 1 already) and define $\hat p_\pm$ accordingly:

$$\hat p_\pm = \tfrac{1}{2}(1\mp\hat\phi_1)(1-\hat\phi_2). \tag{18}$$

It is noted in Lemma 1 that we may equivalently define

$$(p_+,p_-) := \begin{cases}(p,q) & \text{if } \tilde s > 0,\\ (q,p) & \text{if } \tilde s < 0.\end{cases}$$

Recall the definitions $g := \phi_3|\tilde{\mathcal{I}}| = m_2^{-1}\sqrt{4m_1^2m_2+m_3^2}$, $m_1 := r(\phi)\tilde{\mathcal{I}}^2$, $m_2 := r(\phi)\phi_2\tilde{\mathcal{I}}^2$, and $m_3 := r(\phi)\phi_1\phi_2\phi_3\tilde{\mathcal{I}}^3$. Also recall the event Ω_n defined in Proposition 5, and proved therein to satisfy $\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)}\mathbb{P}_\theta(\Omega_n^c) \le 14\exp\left(-\frac{Cn\gamma^*\delta^2\epsilon^4\zeta^6}{L^3+\max(\tau,\sqrt{L})^3\delta\epsilon^2\zeta^3}\right)$ for a constant $C>0$:

$$\Omega_n := \Big\{\max_{j=1,2}\Big|\frac{\hat m_j}{m_j}-1\Big| \le \frac{1}{2},\ \max_{j=1,2,3}|\hat m_j - m_j| \le \frac{gm_2}{44\max(1,g)}\Big\}.$$

Its definition is according to the needs of the proof of Theorem 3 which are more stringent than those of the current result. In particular, note that on Ω_n we have $\max_{j=1,2,3}|\hat m_j - m_j| \le$

$\frac{|\tilde{\mathcal{I}}|^3 r(\phi)\phi_2\phi_3}{20\max(|\phi_1|,(1-\phi_1^2)\phi_3|\tilde{\mathcal{I}}|)}$, as a consequence of the fact that $|\phi_1| \le 1$; this latter bound is what we will use for the current theorem.

We decompose

$$\mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2\Big) = \mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2 \mathbf{1}_{\Omega_n^c}\Big) + \mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2 \mathbf{1}_{\Omega_n}\Big)$$
$$\le \mathbb{P}_\theta(\Omega_n^c) + \mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2 \mathbf{1}_{\Omega_n}\Big),$$

We have

$$\hat{p}_\pm - p_\pm = -\frac{1}{2}(\hat{\phi}_2 - \phi_2) \mp \frac{1}{2}(\hat{\phi}_1 - \tilde{s}\phi_1) \pm \frac{\tilde{s}\phi_1}{2}(\hat{\phi}_2 - \phi_2) \mp \frac{\hat{\phi}_2}{2}(\hat{\phi}_1 - \tilde{s}\phi_1),$$

hence, using that $|\hat{\phi}_2| \le 1$ and $|\phi_1| \le 1$,

$$|\hat{p}_\pm - p_\pm| \le |\hat{\phi}_1 - \phi_1| + |\hat{\phi}_2 - \phi_2|.$$

Using Lemmas 12 and 14 below and Proposition 4, we get for a constant K

$$\mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2 \mathbf{1}_{\Omega_n}\Big) \le 2\mathbb{E}_\theta\Big(|\hat{\phi}_1 - \tilde{s}\phi_1|^2 \mathbf{1}_{\Omega_n}\Big) + 2\mathbb{E}_\theta\Big(|\hat{\phi}_2 - \phi_2|^2 \mathbf{1}_{\Omega_n}\Big)$$
$$\le 2\Big(\frac{53^2\max(1,g^2)}{\phi_2^4\phi_3^6|\tilde{\mathcal{I}}|^6} + \frac{16}{m_1^2}\Big)\mathbb{E}_\theta\Big(\max_{j=1,2,3}|\hat{m}_j - m_j|^2\Big)$$
$$\le 2K\Big(\frac{53^2\max(1,g^2)}{\phi_2^4\phi_3^6|\tilde{\mathcal{I}}|^6} + \frac{16}{m_1^2}\Big)\Big(\frac{L^3}{n\gamma^*} + \frac{\max(\tau,\sqrt{L})^6}{(n\gamma^*)^2}\Big).$$

Therefore, there is a universal constant $B \ge 1$ such that

$$\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)} \mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2 \mathbf{1}_{\Omega_n}\Big)$$
$$\le \frac{BL^3\max(\delta^2,\epsilon^2\zeta^2)}{\delta^2\epsilon^4\zeta^6}\frac{1}{n\gamma^*} + \frac{B\max(\tau,L)^6\max(\delta^2,\epsilon^2\zeta^2)}{\delta^2\epsilon^4\zeta^6}\frac{1}{(n\gamma^*)^2}$$
$$\le \frac{2BL^3\max(\delta^2,\epsilon^2\zeta^2)}{\delta^2\epsilon^4\zeta^6}\frac{1}{n\gamma^*},$$

since $L \ge 1$ and $\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)} \mathbb{E}_\theta\Big(|\hat{p}_\pm - p_\pm|^2 \mathbf{1}_{\Omega_n}\Big) \le 1$. Lemmas 12 and 14 therefore conclude the proof.

LEMMA 12. *Suppose*

$$\max_{j=1,2}\Big|\frac{\hat{m}_j}{m_j} - 1\Big| \le \frac{1}{2}, \quad \text{and,} \quad \max_{j=1,2,3}|\hat{m}_j - m_j| \le \frac{|\tilde{\mathcal{I}}|^3 r(\phi)\phi_2\phi_3}{20\max(|\phi_1|,(1-\phi_1^2)\phi_3|\tilde{\mathcal{I}}|)}.$$

Then,

$$|\hat{\phi}_1 - \tilde{s}\phi_1| \le \frac{53\max(1,\phi_3|\tilde{\mathcal{I}}|)}{\phi_2^2\phi_3^3|\tilde{\mathcal{I}}|^3}\max_{j=1,2,3}|\hat{m}_j - m_j|.$$

PROOF. We use the notations $\Delta_1 = \hat{m}_1 - m_1$, $\Delta_2 = (\hat{m}_2)_+ - m_2$, and $\Delta_3 = \hat{m}_3 - m_3$. Then, we define

$$\hat{v} := 4\hat{m}_1^2(\hat{m}_2)_+ + \hat{m}_3^2,$$

$$v := 4m_1^2 m_2 + m_3^2,$$
$$h := \hat{v} - v,$$
$$\xi := 8m_1 m_2 \Delta_1 + 4m_1^2 \Delta_2 + 8m_1 \Delta_1 \Delta_2 + 4m_2 \Delta_1^2 + 4\Delta_1^2 \Delta_2,$$
$$\eta := 2m_3 \Delta_3 + \Delta_3^2.$$

Lemma 13 below tells us that $10\max(|\phi_1|, (1-\phi_1^2)\phi_3|\tilde{\mathcal{I}}|)|r(\phi)\phi_2\phi_3\tilde{\mathcal{I}}^3|\max_{j=1,2,3}|\Delta_j|$.

Furthermore, it is seen that $\sqrt{v} = |\tilde{\mathcal{I}}|^3 r(\phi)\phi_2\phi_3 = |\tilde{\mathcal{I}}|m_2\phi_3$ (see Lemma 4) and then under the conditions of this lemma, we have $|h| \leq v/2$ and $|\Delta_3| \leq (1/2)|m_3| = (1/2)\phi_1\phi_3\tilde{\mathcal{I}}|m_2| \leq \sqrt{v}/2$. Consequently, $1 - \frac{\Delta_3^2}{(\sqrt{v+h}+\sqrt{v})^2} \geq 3/4$ and $(v+h)^{1/2}[(v+h)^{1/2} + v^{1/2}] \geq (1+\sqrt{2})v/2 \geq v$ and hence

$$\begin{aligned}
|\hat{\phi}_1 - \tilde{s}\phi_1| &\leq \frac{|\phi_1\xi|}{v} + \frac{4}{3v}\Big[2|\Delta_3|(1-\phi_1^2)v^{1/2} + |\phi_1|\Delta_3^2|\xi|v^{-1} + |\Delta_3\xi|v^{-1/2}\Big] \\
&\leq \frac{28}{v}m_1^2 \max_{j=1,2}|\Delta_j| + \frac{8}{3}(1-\phi_1^2)v^{-1/2}|\Delta_3| + \frac{4}{3}|\xi|[1/2 + |\phi_1|/4] \\
&\leq 28\frac{m_1^2}{v}\max_{j=1,2}|\Delta_j| + \frac{8}{3}(1-\phi_1^2)v^{-1/2}|\Delta_3| + 56\frac{m_1^2}{v}\max_{j=1,2}|\Delta_j| \\
&\leq 42(\phi_2^2\phi_3^2\tilde{\mathcal{I}}^2)^{-1}\max_{j=1,2}|\Delta_j| + \frac{32}{3}(\phi_2^2\phi_3^3\tilde{\mathcal{I}}^3)^{-1}|\Delta_3| \\
&\leq 53(\phi_2^2\phi_3^3\tilde{\mathcal{I}}^3)^{-1}\max(\phi_3\tilde{\mathcal{I}}, 1)\max_{j=1,2,3}|\Delta_j|.
\end{aligned}$$

The conclusion follows since $x \mapsto (x)_+$ is 1-Lipschitz and thus $|\Delta_2| = |(\hat{m}_2)_+ - m_2| = |(\hat{m}_2)_+ - (m_2)_+| \leq |\hat{m}_2 - m_2|$, so that $\max_{j=1,2,3}|\Delta_j| \leq \max_{j=1,2,3}|\hat{m}_j - m_j|$. □

LEMMA 13. *Define $v = 4m_1^2 m_2 + m_3^2$, $\hat{v} = 4\hat{m}_1^2(\hat{m}_2)_+ + \hat{m}_3^2$. Then*

$$|\hat{v} - v| \leq 10\max(|\phi_1|, (1-\phi_1^2)\phi_3|\tilde{\mathcal{I}}|)|r(\phi)\phi_2\phi_3\tilde{\mathcal{I}}^3|\max_{j=1,2,3}|\Delta_j|,$$

where $\Delta_j = \hat{m}_j - m_j$, $j = 1, 3$ and $\Delta_2 = (\hat{m}_2)_+ - m_2$.

PROOF. Define

$$h := \hat{v} - v,$$
$$\xi := 8m_1 m_2 \Delta_1 + 4m_1^2 \Delta_2 + 8m_1 \Delta_1 \Delta_2 + 4m_2 \Delta_1^2 + 4\Delta_1^2 \Delta_2,$$
$$\eta := 2m_3 \Delta_3 + \Delta_3^2.$$

Note that $h = \xi + \eta$. By mimicking the proof of [2] Proposition 3, it is found that

$$\hat{\phi}_1 - \tilde{s}\phi_1 = \frac{\phi_1\xi + \frac{-2\Delta_3(1-\phi_1^2)v^{1/2} + \frac{\phi_1\Delta_3^2\xi}{((v+h)^{1/2}+v^{1/2})^2} - \frac{\Delta_3\xi}{(v+h)^{1/2}+v^{1/2}}}{1-\Delta_3^2/((v+h)^{1/2}+v^{1/2})^2}}{(v+h)^{1/2}[(v+h)^{1/2}+v^{1/2}]}$$

We note that the assumptions of the lemma imply that $|\Delta_j| \leq |m_j|$ for $j = 1, 2, 3$; recall also that $0 \leq m_2 = m_1 \leq |m_1|$. Thus,

$$\begin{aligned}
|\xi| &= \Big|8m_1 m_2 \Delta_1 + 4m_1^2 \Delta_2 + 8m_1 \Delta_1 \Delta_2 + 4m_2 \Delta_1^2 + 4\Delta_1^2 \Delta_2\Big| \\
&\leq 28m_1^2 \max_{j=1,2}|\Delta_j|.
\end{aligned}$$

Since $|\eta| \leq 2|m_3\Delta_3| + \Delta_3^2 \leq 3|m_3\Delta_3|$, it also follows that (recall $m_1 = r(\phi)\tilde{\mathcal{I}}^2$, $m_3 = \phi_1\phi_2\phi_3 r(\phi)\tilde{\mathcal{I}}^3$, $r(\phi) = (1/4)(1-\phi_1^2)\phi_2\phi_3^2$)

$$
\begin{aligned}
|h| &\leq \left(28m_1^2 + 3|m_3|\right) \max_{j=1,2,3} |\Delta_j| \\
&= |r(\phi)\phi_2\phi_3\tilde{\mathcal{I}}^3| \Big(3|\phi_1| + \frac{28|r(\phi)\tilde{\mathcal{I}}|}{|\phi_2\phi_3|}\Big) \max_{j=1,2,3} |\Delta_j| \\
&= |r(\phi)\phi_2\phi_3\tilde{\mathcal{I}}^3| \Big(3|\phi_1| + 7(1-\phi_1^2)\phi_3|\tilde{\mathcal{I}}|\Big) \max_{j=1,2,3} |\Delta_j| \\
&\leq 10 \max\left(|\phi_1|, (1-\phi_1^2)\phi_3|\tilde{\mathcal{I}}|\right) |r(\phi)\phi_2\phi_3\tilde{\mathcal{I}}^3| \max_{j=1,2,3} |\Delta_j|.
\end{aligned}
$$

This concludes the proof. □

LEMMA 14. *The following bounds holds true.*

$$
|\hat{\phi}_2 - \phi_2| \leq 2 \min\left(1, \frac{2\max_{j=1,2}|\hat{m}_j - m_j|}{|m_1|}\right).
$$

PROOF. We let $\Delta_1 := \hat{m}_1 - m_1$ and $\Delta_2 := \hat{m}_2 - m_2$. We also let $f(x) := \max(-1, \min(x, 1))$. It is easily seen that $|f(x) - f(y)| \leq \min(2, |x-y|)$. Suppose first that $|\Delta_1| > |m_1|/2$. Then, $|\hat{\phi}_2 - \phi_2| \leq 2 \leq \min(2, \frac{4|\Delta_1|}{|m_1|})$. On the other hand, if $|\Delta_1| \leq |m_1|/2$, then, recalling that $m_2 \leq |m_1|$ we have

$$
\begin{aligned}
|\hat{\phi}_2 - \phi_2| &= |f(\hat{m}_2/\hat{m}_1) - f(m_2/m_1)| \\
&\leq \min\Big(2, \Big|\frac{m_2 + \Delta_2}{m_1 + \Delta_1} - \frac{m_2}{m_1}\Big|\Big) \\
&= \min\Big(2, \Big|\frac{m_1\Delta_2 - m_2\Delta_1}{m_1(m_1 + \Delta_1)}\Big|\Big) \\
&\leq \min\Big(2, \frac{2|\Delta_1| + 2|\Delta_2|}{|m_1|}\Big).
\end{aligned}
$$

The conclusion follows since $x \mapsto (x)_+$ is 1-Lipschitz and thus $|\Delta_2| = |(\hat{m}_2)_+ - m_2| = |(\hat{m}_2)_+ - (m_2)_+| \leq |\hat{m}_2 - m_2|$. □

A.4. Proof of Theorem 3. As in Appendix A.3, rather than allow an arbitrary permutation to account for the label-switching, we give a specific (unobserved) permutation. We recall the definitions of the estimators of f_0 and f_1 from Section 4, here writing as $\check{f}_\pm$ to align with notation used in Lemma 1. We define

$$
\begin{aligned}
g &:= \phi_3|\tilde{\mathcal{I}}| = \frac{\sqrt{4m_1^2 m_2 + m_3^2}}{m_2}, \qquad G := \frac{m_1\psi_2}{\tilde{\mathcal{I}}}, \\
f_\pm &:= \psi_1 \pm \frac{g(1 \mp \tilde{s}\phi_1)}{2m_1} G,
\end{aligned}
$$

and

$$
\hat{g} := \frac{\sqrt{4\hat{m}_1^2(\hat{m}_2)_+ + \hat{m}_3^2}}{\hat{m}_2} \mathbf{1}_{\{\hat{m}_2 > 0\}}, \qquad \hat{G}^{\Phi_{Jk}} := \mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \Phi_{Jk}) - \mathbb{P}_n^{(1)}(\tilde{\psi}_2)\mathbb{P}_n^{(1)}(\Phi_{Jk}),
$$

$$
\hat{f}_\pm^{\Phi_{Jk}} := \mathbb{P}_n^{(1)}(\Phi_{Jk}) + \frac{\hat{g}(1 - \hat{\phi}_1)}{2\hat{m}_1} \mathbf{1}_{\{\hat{m}_1 \neq 0\}} \hat{G}^{\Phi_{Jk}}.
$$

Then, defining $\hat f_\pm^{\Psi_{jk}}$ and $\hat G^{\Psi_{jk}}$ correspondingly we set

$$\hat f_\pm := \sum_{k=0}^{2^J-1} \hat f_\pm^{\Phi_{Jk}} \Phi_{Jk} + \sum_{j=J}^{J_n-1}\sum_{k=0}^{2^j-1} \hat f_\pm^{\Psi_{jk}}\Psi_{jk} + \sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \left(\sum_{k\in\mathfrak{B}_{j\ell}} \hat f_\pm^{\Psi_{jk}}\Psi_{jk}\right)\mathbf{1}_{\{\|\hat f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma\hat S_n\}},$$

$$\check f_\pm := \max\big(0, \min\big(\check T, \hat f_\pm\big)\big),$$

where $J_n := \inf\{j\ge J : 2^j \ge \log(n)\}$, $N = 2^{J_n}$, and $\mathfrak{B}_{j\ell} := \{k : (\ell-1)N \le k \le \ell N - 1\}$ and $\tilde{\jmath}_n$ is the largest integer such that $2^{\tilde{\jmath}_n} \le \frac{n}{\log(n)\tau^2}$ (recall we assume that $\tilde{\jmath}_n$ is larger than J_n) and where $\|\hat f_\pm^{\mathfrak{B}_{j\ell}}\|^2 := \sum_{k\in\mathfrak{B}_{j\ell}}(\hat f_\pm^{\Psi_{jk}})^2$, $\Gamma>0$ is a tuning parameter, and

$$\hat S_n := \sqrt{\frac{\log(n)}{n}}\max\Big(1, \frac{\hat g}{|\hat m_1|}\Big)\mathbf{1}_{\{\hat m_1\neq 0\}}.$$

Recall the event $\Omega_n = \Big\{\max_{j=1,2}\Big|\frac{\hat m_j}{m_j}-1\Big| \le \frac12,\ \max_{j=1,2,3}|\hat m_j - m_j| \le \frac{g m_2}{44\max(1,g)}\Big\}$ defined in Proposition 5 which by the proposition satisfies for a universal constant $C>0$

$$\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)} \mathbb{P}_\theta(\Omega_n^c) \le 7\exp\left(-\frac{Cn\gamma^*\delta^2\epsilon^4\zeta^6}{L^3+\max(\tau,\sqrt{L})^3\delta\epsilon^2\zeta^3}\right).$$

Decompose

$$\begin{aligned}\mathbb{E}_\theta\Big(\|\check f_\pm - f_\pm\|^2_{L^2}\Big) &= \mathbb{E}_\theta\Big(\|\check f_\pm - f_\pm\|^2_{L^2}\mathbf{1}_{\Omega_n^c}\Big) + \mathbb{E}_\theta\Big(\|\check f_\pm - f_\pm\|^2_{L^2}\mathbf{1}_{\Omega_n}\Big)\\ &\le \check T^2\mathbb{P}_\theta(\Omega_n^c) + \mathbb{E}_\theta\Big(\|\hat f_\pm - f_\pm\|^2_{L^2}\mathbf{1}_{\Omega_n}\Big)\end{aligned}$$

where the last line follows because $0\le f_\pm, \check f_\pm \le \check T$ since $\check T \ge L$ by assumption, and because $|\check f_\pm - f_\pm| \le |\hat f_\pm - f_\pm|$ pointwise. The first term is included in the theorem and it remains to bound the second term. We decompose as follows (recall that $\tilde{\jmath}_n > J_n$ by assumption and the sum over ℓ is the sum over blocks from $\ell = 0$ to $\ell = 2^j/N - 1$)

$$\begin{aligned}\mathbb{E}_\theta\Big(\|\hat f_\pm - f_\pm\|^2_{L^2}\mathbf{1}_{\Omega_n}\Big) &= \mathbb{E}_\theta\Big(\|\hat f_\pm^{J_n} - f_\pm^{J_n}\|^2_{L^2}\mathbf{1}_{\Omega_n}\Big)\\ &\quad+\mathbb{E}_\theta\left(\sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat f_\pm^{\mathfrak{B}_{j\ell}}\|\le\Gamma\hat S_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|\le 2\Gamma\hat S_n\}}\mathbf{1}_{\Omega_n}\right)\\ &\quad+\mathbb{E}_\theta\left(\sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat f_\pm^{\mathfrak{B}_{j\ell}}\|\le\Gamma\hat S_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|> 2\Gamma\hat S_n\}}\mathbf{1}_{\Omega_n}\right)\\ &\quad+\mathbb{E}_\theta\left(\sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \|\hat f_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma\hat S_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|\le \frac12\Gamma\hat S_n\}\}}\mathbf{1}_{\Omega_n}\right)\\ &\quad+\mathbb{E}_\theta\left(\sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \|\hat f_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma\hat S_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|> \frac12\Gamma\hat S_n\}}\mathbf{1}_{\Omega_n}\right)\\ &\quad+\mathbb{P}_\theta(\Omega_n)\sum_{j>\tilde{\jmath}_n}\sum_{k=0}^{2^j-1}|f_\pm^{\Psi_{jk}}|^2\end{aligned}$$

where we have used the convention that for any function f the notation f^{J_n} stands for the projection $f_\pm^{J_n} := \sum_{k=0}^{2^J-1} f_\pm^{\Phi_{Jk}}\Phi_{Jk} + \sum_{j=J}^{J_n-1}\sum_{k=0}^{2^j-1} f_\pm^{\Psi_{jk}}\Psi_{jk}$. Recall that $f^{\mathfrak{B}_{j\ell}}$ denotes the vector of coefficients $(\langle f, \Psi_{jk}\rangle : (j,k) \in \mathfrak{B}_{j\ell})$ and $\|\cdot\|$ the euclidean norm. We call the terms in the previous decomposition $R_1(\theta), \dots, R_6(\theta)$, respectively. To ease the notations in the proof, we also introduce the quantities

$$\hat{\omega}_\pm \coloneqq \pm \frac{\hat{g}(1 \mp \hat{\phi}_1)}{\hat{m}_1} \mathbf{1}_{\{\hat{m}_1 \neq 0\}}, \qquad \omega_\pm \coloneqq \pm \frac{g(1 \mp \tilde{s}\phi_1)}{m_1} \tag{19}$$

and

$$S_n \coloneqq \sqrt{\frac{\log(n)}{n}} \max\Big(1, \frac{g}{|m_1|}\Big). \tag{20}$$

In the next subsections we prove the following bounds, uniformly over $\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R) \cap \Sigma_{\gamma^*}(L)$:

$$R_1(\theta) \le \frac{BL^2}{\delta^2\epsilon^2\zeta^2}\frac{\log(n)}{n\gamma^*} + \frac{BL^3}{\delta^2\epsilon^4\zeta^4}\frac{1}{n\gamma^*} + \frac{B\max(\tau,\sqrt{L})^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2}$$

$$R_2(\theta) \le \frac{BR^2}{\min(1,s_\pm)}\Big(\frac{\Gamma^2}{R^2\delta^2\epsilon^2\zeta^2 n}\Big)^{2s_\pm/(2s_\pm+1)} + \frac{BR^2}{\min(1,s_\pm)}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_\pm}$$

$$R_3(\theta) \le \frac{BL^3}{\delta^2\epsilon^4\zeta^4}\frac{1}{n\gamma^*} + \frac{B\max(\tau,\sqrt{L})^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2},$$

$$R_4(\theta) \le \frac{BL^3}{\delta^2\epsilon^4\zeta^4}\frac{1}{n\gamma^*} + \frac{B\max(\tau,\sqrt{L})^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2},$$

$$\begin{aligned} R_5(\theta) \le{}& \frac{BL^2}{\Gamma^2\gamma^*}\Bigg(\frac{R^2}{\min(1,s_\pm)}\Big(\frac{\Gamma^2}{R^2\delta^2\epsilon^2\zeta^2 n}\Big)^{2s_\pm/(2s_\pm+1)} + \frac{R^2}{\min(1,s_\pm)}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_\pm}\Bigg) \\ &+ \frac{BL^3}{\delta^2\epsilon^4\zeta^6}\frac{1}{n\gamma^*} + \frac{B\max(\tau,\sqrt{L})^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2}, \end{aligned}$$

$$R_6(\theta) \le \frac{BR^2}{\min(1,s_\pm)}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_\pm}.$$

Combining will yield the theorem.

A.4.1. *Control of* R_1. Using Lemma 15 to control $\|\hat{f}^{J_n}_\pm - f^{J_n}_\pm\|_{L^2}$ and Proposition 10 in Section A.4.7 to control $|\hat{\omega}_\pm - \omega_\pm|$, the bounds $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ and $\|G^{J_n}\|_{L^2} = |m_1|\|\psi^{J_n}_2\|_{L^2}/|\tilde{\mathcal{I}}| \le (8/7)|m_1|$ allow us to deduce

$$\begin{aligned} R_1(\theta) &\coloneqq \mathbb{E}_\theta\Big(\|\hat{f}^{J_n}_\pm - f^{J_n}_\pm\|^2_{L^2}\mathbf{1}_{\Omega_n}\Big) \\ &\le 3\mathbb{E}_\theta\Big(\|\hat{\psi}^{J_n}_1 - \psi^{J_n}_1\|^2_{L^2}\Big) + \frac{12g^2}{m_1^2}\mathbb{E}_\theta\Big(\|\hat{G}^{J_n} - G^{J_n}\|^2_{L^2}\Big) + \frac{3\|G^{J_n}\|^2_{L^2}}{4}\mathbb{E}_\theta\Big(|\hat{\omega}_\pm - \omega_\pm|^2\mathbf{1}_{\Omega_n}\Big). \end{aligned}$$

ie.

$$\begin{aligned} R_1(\theta) \le{}& 3\mathbb{E}_\theta\Big(\|\hat{\psi}^{J_n}_1 - \psi^{J_n}_1\|^2_{L^2}\Big) + \frac{12g^2}{m_1^2}\mathbb{E}_\theta\Big(\|\hat{G}^{J_n} - G^{J_n}\|^2_{L^2}\Big) \\ &+ \frac{3\cdot 8^2\cdot 83^2\max(1,\phi_3^2\tilde{\mathcal{I}}^2)}{4\cdot 7^2 m_2^2}\mathbb{E}_\theta\Big(\max_{j=1,2,3}|\hat{m}_j - m_j|^2\Big). \end{aligned} \tag{21}$$

Proposition 6 tells us that

$$\mathbb{P}_\theta\left(\|\hat{\psi}_1^{J_n} - \psi_1^{J_n}\|_{L^2} \geq C\sqrt{\frac{Lx}{n\gamma^*}} + C 2^{J_n/2}\frac{x}{n\gamma^*} \right) \leq 24^{2^{J_n}} e^{-x},$$

hence, using that $2^{J_n} \leq 2\log(n)$ for $n \geq 2$, for a sufficient large constant $\alpha > 0$ we may apply Lemma 8 with $a = C\sqrt{2\log(n)}/\gamma^*$, $b = C\sqrt{L/\gamma^*}$, $c = 24^{2\log(n)}$ and $d^2 = \alpha C^2 L \log(n)/(n\gamma^*)$

$$\begin{aligned}
&\mathbb{E}_\theta\Big(\|\hat{\psi}_1^{J_n} - \psi_1^{J_n}\|_{L^2}^2 \Big) \\
&\quad \leq \alpha C^2 L \frac{\log(n)}{n\gamma^*} + \mathbb{E}_\theta\Big(\|\hat{\psi}_1^{J_n} - \psi_1^{J_n}\|_{L^2}^2 \mathbf{1}_{\{\|\hat{\psi}_1^{J_n} - \psi_1^{J_n}\|_{L^2}^2 > \alpha C^2 L\log(n)/(n\gamma^*)\}} \Big) \\
&\quad \leq \alpha C^2 L \frac{\log(n)}{n\gamma^*} + c\Big(d^2 + \frac{5b^2}{2n} + \frac{7a^2}{2n^2} \Big) e^{-nd^2/(2b^2+8ad)} \\
&\quad \leq \alpha C^2 L \frac{\log(n)}{n\gamma^*} + C^2 24^{2\log(n)} \Big(\frac{\alpha L\log(n)}{n\gamma^*} + \frac{5L}{2n\gamma^*} + \frac{14\log(n)}{2(n\gamma^*)^2} \Big) e^{-nd^2/(2b^2+8ad)} \\
&\quad \leq \alpha C^2 L \frac{\log(n)}{n\gamma^*} + C^2 24^{2\log(n)} \Big(\alpha L + \frac{5L}{2} + 7 \Big) \log(n) e^{-nd^2/(2b^2+8ad)}
\end{aligned}$$

where the last line follows because $n\gamma^* \geq \tau^3 \geq 1$. Let us now study the argument of the exponential in the last display. If $2b^2 \geq 8ad$, then

$$\frac{nd^2}{2b^2+8ad} \geq \frac{nd^2}{4b^2} = \frac{\alpha}{4}\log(n),$$

while if $2b^2 < 8ad$, then

$$\frac{nd^2}{2b^2+8ad} \geq \frac{nd^2}{16ad} = \frac{n\gamma^*\sqrt{\alpha C^2 L\log(n)/(n\gamma^*)}}{16C\sqrt{2\log n}} \geq \frac{\sqrt{\alpha L}}{16\sqrt{2}}\sqrt{n\gamma^*} \geq \frac{\sqrt{\alpha}}{16\sqrt{2}}\log(n)$$

because by assumption $n\gamma^* \geq \frac{\log(n)^2}{L}$. Hence, since $L \leq n$ and $\gamma^* \leq 1$ it is possible to choose $\alpha > 0$ universally such that

$$\mathbb{E}_\theta\Big(\|\hat{\psi}_1^{J_n} - \psi_1^{J_n}\|_{L^2}^2 \Big) \leq 2\alpha C^2 L \frac{\log(n)}{n\gamma^*}.$$

Similarly, Proposition 7 tells us that

$$\mathbb{P}_\theta\left(\|\hat{G}^{J_n} - G^{J_n}\|_{L^2} \geq CL\sqrt{\frac{x}{n\gamma^*}} + C\max(\tau 2^{J_n/2}, \sqrt{L}2^{J_n/2}, \tau\sqrt{L})\frac{x}{n\gamma^*} \right) \leq 4\cdot 24^{2^{J_n}} e^{-x},$$

hence, for any $\alpha > 0$, using that $2^{J_n} \leq 2\log(n)$ for $n \geq 2$, Lemma 8 with $a = C\tau\sqrt{2L\log(n)}/\gamma^*$, $b = CL/\sqrt{\gamma^*}$, $c = 4 \times 24^{2\log n}$, and $d^2 = \alpha C^2 L^2 \log(n)/(n\gamma^*)$ [and by remarking that $\max(\tau 2^{J_n/2}, \sqrt{L}2^{J_n/2}, \tau\sqrt{L}) \leq \tau\sqrt{L}2^{J_n/2}$] yields

$$\begin{aligned}
&\mathbb{E}_\theta\Big(\|\hat{G}^{J_n} - G^{J_n}\|_{L^2}^2 \Big) \\
&\quad \leq \alpha C^2 L^2 \frac{\log(n)}{n\gamma^*} + c\Big(d^2 + \frac{5b^2}{2n} + \frac{7a^2}{2n^2} \Big) e^{-nd^2/(2b^2+8ad)} \\
&\quad \leq \alpha C^2 L^2 \frac{\log(n)}{n\gamma^*} + 4C^2 24^{2\log(n)} \Big(\frac{\alpha L^2\log(n)}{n\gamma^*} + \frac{5L^2}{2n\gamma^*} + \frac{14\tau^2 L\log(n)}{2(n\gamma^*)^2} \Big) e^{-nd^2/(2b^2+8ad)}.
\end{aligned}$$

Let us study the argument of the exponential in the last display. If $2b^2 \geq 8ad$, then

$$\frac{nd^2}{2b^2+8ad} \geq \frac{nd^2}{4b^2} = \frac{\alpha}{4}\log(n)$$

while if $2b^2 < 8ad$, then

$$\frac{nd^2}{2b^2+8ad} \geq \frac{nd^2}{16ad} = \frac{n\gamma^* \sqrt{\alpha C^2 L^2 \log(n)/(n\gamma^*)}}{16C\tau\sqrt{L}2^{J_n/2}} \geq \frac{\sqrt{\alpha L}}{32\tau}\sqrt{n\gamma^*} \geq \frac{\sqrt{\alpha}}{32}\log(n)$$

because by assumption $n\gamma^* \geq \frac{\tau^2\log(n)^2}{L}$. Since by assumption $L \leq n$ and $n\gamma^* \geq \tau^3 \geq 1$, it is possible to choose $\alpha > 0$ universally such that

$$\mathbb{E}_\theta\Big(\|\hat{G}^{J_n} - G^{J_n}\|_{L^2}^2\Big) \leq 2\alpha C^2 L^2 \frac{\log(n)}{n\gamma^*}.$$

Returning to (21) and feeding the bound for $E_\theta \max_j|\hat{m}_j - m_j|^2$ from Proposition 4, we deduce that

$$R_1(\theta) \leq 6\alpha C^2 L\Big(1+\frac{g^2 L}{m_1^2}\Big)\frac{\log(n)}{n\gamma^*} + \frac{3\cdot 83^2\cdot 40C^2L^3\max(1,g^2)}{n\gamma^* m_2^2} + \frac{3\cdot 83^2\cdot 64C^2\max(\tau,\sqrt{L})^6}{(n\gamma^*)^2 m_2^2}.$$

Finally, we remark $\frac{g^2}{m_1^2} \leq \frac{16}{\delta^2\epsilon^2\zeta^2\tilde{\mathcal{I}}^2}$ and $\frac{\max(1,g^2)}{m_2^2} \leq \frac{16}{\delta^2\epsilon^4\zeta^4\tilde{\mathcal{I}}^4}$ by Lemma 4 and by the assumption that $\zeta \leq 1$. Thus, there exists a universal constant $B > 0$ such that

$$\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)} R_1(\theta) \leq \frac{BL^2}{\delta^2\epsilon^2\zeta^2}\frac{\log(n)}{n\gamma^*} + \frac{BL^3}{\delta^2\epsilon^4\zeta^4}\frac{1}{n\gamma^*} + \frac{B\max(\tau,\sqrt{L})^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2}.$$

A.4.2. *Control of* R_2. From equation (10) whenever $\theta \in \Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)$ it is the case that $\sup_{j\geq J} 2^{2js_\pm}\sum_k |f_\pm^{\Psi_{jk}}|^2 \leq R^2$. This in particular implies that $\sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \leq R^2 2^{-2js_\pm}$. Moreover $\hat{S}_n \leq 4S_n$ on Ω_n by Proposition 11 in Section A.4.7. Then, since $J_n \leq \tilde{j}_n$,

$$\begin{aligned}
R_2(\theta) &:= \mathbb{E}_\theta\Bigg(\sum_{j=J_n}^{\tilde{j}_n}\sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}\|\leq\Gamma\hat{S}_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|\leq 2\Gamma\hat{S}_n\}}\mathbf{1}_{\Omega_n}\Bigg) \\
&\leq \sum_{j=J_n}^{\tilde{j}_n}\sum_\ell \min\Big(\|f_\pm^{\mathfrak{B}_{j\ell}}\|^2, 8\Gamma S_n\Big)^2 \\
&\leq \sum_{j=J_n}^{\tilde{j}_n}\min\Bigg(\sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2, \frac{2^j}{N}\cdot 64\Gamma^2 S_n^2\Bigg) \\
&\leq \sum_{j=J_n}^{\tilde{j}_n}\min\Big(R^2 2^{-2js_\pm}, \frac{2^j}{N}\cdot 64\Gamma^2 S_n^2\Big).
\end{aligned}$$

Define $A = \sup\{0 \leq j \leq \tilde{j}_n : 2^{-j(s_\pm+1/2)} > 8\Gamma S_n/(R\sqrt{N})\}$, so that the first term in the minimum is the smaller exactly when $j > A$. Then we observe that $2^A < (R^2N/(64\Gamma^2 S_n^2))^{1/(2s_\pm+1)}$ and $2^{A+1} \geq \min\{(R^2N/(64\Gamma^2S_n^2))^{1/(2s_\pm+1)}, n/(\tau^2\log n)\}$ (for the latter recall that $\tilde{j}$ is the largest integer such that $2^{\tilde{j}} \leq n/(\tau^2\log n)$), and we calculate

$$R_2(\theta) \leq \frac{64\Gamma^2 S_n^2}{N}\sum_{j=0}^{A} 2^j + R^2\sum_{j=A+1}^{\infty} 2^{-2js_\pm}$$

$$\leq \frac{128\Gamma^2 S_n^2}{N}\left(\frac{c^2R^2N}{64\Gamma^2 S_n^2}\right)^{1/(2s_\pm+1)} + \frac{R^2}{1-2^{-2s_\pm}}\max\left(\frac{\tau^2\log(n)}{n}, \left(\frac{64\Gamma^2 S_n^2}{R^2N}\right)^{1/(2s_\pm+1)}\right)^{2s_\pm}$$

$$= 2R^2\left(\frac{64\Gamma^2 S_n^2}{R^2N}\right)^{2s_\pm/(2s_\pm+1)} + \frac{R^2}{1-2^{-2s_\pm}}\max\left(\frac{\tau^2\log(n)}{n}, \left(\frac{64\Gamma^2 S_n^2}{R^2N}\right)^{1/(2s_\pm+1)}\right)^{2s_\pm}.$$

Recalling that $S_n = \sqrt{(\log n)/n}\max(1, g/|m_1|)$ and $N > \log n$, we deduce that

$$R_2(\theta) \leq 2R^2\left(\frac{64\Gamma^2\max(1, g^2/m_1^2)}{R^2n}\right)^{2s_\pm/(2s_\pm+1)}$$
$$+ \frac{R^2}{1-2^{-2s_\pm}}\max\left(\frac{\tau^2\log(n)}{n}, \left(\frac{64\Gamma^2\max(1, g^2/m_1^2)}{R^2n}\right)^{1/(2s_\pm+1)}\right)^{2s_\pm}.$$

Hence, recalling that $|\tilde{\mathcal{I}}| \geq 7/8$ and the result of Lemma 4, there exists a universal constant $B > 0$ such that

$$\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)} R_2(\theta) \leq \frac{BR^2}{\min(1,s_\pm)}\left(\frac{\Gamma^2}{R^2\delta^2\epsilon^2\zeta^2 n}\right)^{2s_\pm/(2s_\pm+1)} + \frac{BR^2}{\min(1,s_\pm)}\left(\frac{\tau^2\log(n)}{n}\right)^{2s_\pm}$$

A.4.3. *Control of* R_3. We remark that on the event $\{\|\hat{f}^{\mathfrak{B}_{j\ell}}\| \leq \Gamma\hat{S}_n\}\cap\{\|f^{\mathfrak{B}_{j\ell}}\| > 2\Gamma\hat{S}_n\}$ it must that

$$\|f_\pm^{\mathfrak{B}_{j\ell}}\| \leq \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\| + \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}\| \leq \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\| + \frac{1}{2}\|f_\pm^{\mathfrak{B}_{j\ell}}\|$$

and thus $\|f_\pm^{\mathfrak{B}_{j\ell}}\| \leq 2\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|$. Then, since $\frac{1}{4}S_n \leq \hat{S}_n \leq 4S_n$ on the event Ω_n by Proposition 11 in Section A.4.7,

$$R_3(\theta) := \mathbb{E}_\theta\left(\sum_{j=J_n}^{\tilde{J}_n}\sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}\|\leq\Gamma\hat{S}_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|>2\Gamma\hat{S}_n\}}\mathbf{1}_{\Omega_n}\right)$$

$$\leq 4\sum_{j=J_n}^{\tilde{J}_n}\sum_\ell \mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}\|\leq\Gamma\hat{S}_n\}}\mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\|>2\Gamma\hat{S}_n\}}\mathbf{1}_{\Omega_n}\Big)$$

$$\leq 4\sum_{j=J_n}^{\tilde{J}_n}\sum_\ell \mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}-f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma S_n/4\}}\mathbf{1}_{\Omega_n}\Big).$$

Recalling that $\hat{f}_\pm = \hat{\psi}_1 + \frac{1}{2}\hat{\omega}_\pm\hat{G}$, we define $W_1^{\mathfrak{B}_{j\ell}} := \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\|$, $W_2^{\mathfrak{B}_{j\ell}} := \frac{4g}{|m_1|}\|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\|$, and $W_3^{\mathfrak{B}_{j\ell}} := \frac{1}{2}|\hat{\omega}_\pm - \omega_\pm|\|G^{\mathfrak{B}_{j\ell}}\|$, so that a direct calculation (see Lemma 15) yields $\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|_{L^2} \leq W_1^{\mathfrak{B}_{j\ell}} + W_2^{\mathfrak{B}_{j\ell}} + W_3^{\mathfrak{B}_{j\ell}}$. We then observe, writing $\bar{W}^{\mathfrak{B}_{j\ell}} = \max(W_1^{\mathfrak{B}_{j\ell}}, W_2^{\mathfrak{B}_{j\ell}}, W_3^{\mathfrak{B}_{j\ell}})$, that

$$R_3(\theta) \leq 4\sum_{j=J_n}^{\tilde{J}_n}\sum_\ell \mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}-f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma S_n/4\}}\mathbf{1}_{\{\bar{W}^{\mathfrak{B}_{j\ell}}=W_1^{\mathfrak{B}_{j\ell}}\}}\mathbf{1}_{\Omega_n}\Big)$$

$$+ 4\sum_{j=J_n}^{\tilde{J}_n}\sum_\ell \mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}-f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma S_n/4\}}\mathbf{1}_{\{\bar{W}^{\mathfrak{B}_{j\ell}}=W_2^{\mathfrak{B}_{j\ell}}\}}\mathbf{1}_{\Omega_n}\Big)$$

$$+ 4\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}-f_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma S_n/4\}}\mathbf{1}_{\{\bar{W}^{\mathfrak{B}_{j\ell}}=W_3^{\mathfrak{B}_{j\ell}}\}}\mathbf{1}_{\Omega_n}\Big)$$

We call these terms $R_{3,1}$, $R_{3,2}$, and $R_{3,3}$, respectively. Let us start with $R_{3,1}$. Observe that on the event $\Omega_n\cap\{\bar{W}^{\mathfrak{B}_{j\ell}} = W_1^{\mathfrak{B}_{j\ell}}\}$ we have $\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\| \le 3W_1^{\mathfrak{B}_{j\ell}}$. Therefore,

$$R_{3,1} \le 36\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\Big(\big(W_1^{\mathfrak{B}_{j\ell}}\big)^2\mathbf{1}_{\{W_1^{\mathfrak{B}_{j\ell}}>\Gamma S_n/12\}}\Big)$$

Proposition 8 in Section A.4.7 tells us that, for $n\gamma^*\ge 1/99$, there is a universal constant $C>0$ such that for all $\theta\in\Sigma_{\gamma^*}(L)$ and all $x\ge 0$

$$\mathbb{P}_\theta\left(\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| \ge C\sqrt{\frac{Lx}{n\gamma^*}} + C2^{j/2}\frac{x}{n\gamma^*}\right) \le 24^N e^{-x}.$$

Then by Lemma 8 with $a = C2^{j/2}/\gamma^*$, $b = C\sqrt{L/\gamma^*}$, $c = 24^N \le 24^{2\log(n)}$ [$n\ge 2$ so $N\le 2\log(n)$], we find that

$$R_{3,1} \le 36\cdot 24^N\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\Big(\frac{\Gamma^2S_n^2}{144} + \frac{5C^2L}{2n\gamma^*} + \frac{7C^22^j}{2(n\gamma^*)^2}\Big)\exp\Big(-\frac{n\gamma^*\Gamma^2S_n^2/144}{2C^2L+8C2^{j/2}\Gamma S_n/12}\Big).$$

$$\le 36\cdot 24^N\Big(\frac{\Gamma^2\max(1,g^2/m_1^2)}{144} + 5C^2Ln + \frac{14C^2n^2}{2}\Big)\exp\Big(-\frac{n\gamma^*\Gamma^2S_n^2/144}{2C^2L+8C2^{j/2}\Gamma S_n/12}\Big)$$

where the last line follows since there are $2^j/N\le 2^j$ blocks at each level j, and because $2^{\tilde{j}_n}\le n$ by construction whenever $n\ge 3$, and because $n\gamma^*\ge\tau^3\ge 1$. Let us analyse the argument of the exponential in the last display. Firstly if $8C2^{j/2}\Gamma S_n/12 \le 2C^2L$, it is the case that

$$\frac{n\gamma^*\Gamma^2S_n^2/144}{2C^2L+8C2^{j/2}\Gamma S_n/12} \ge \frac{n\gamma^*\Gamma^2S_n^2}{576C^2L} \ge \frac{\gamma^*\Gamma^2}{576C^2L}\log(n)$$

since $S_n = \sqrt{\log(n)/n}\max(1,g/|m_1|)$. Secondly, if $8C2^{j/2}\Gamma S_n/12 > 2C^2L$, it is the case that for any $j\le\tilde{j}_n$

$$\frac{n\gamma^*\Gamma^2S_n^2/144}{2C^2L+8C2^{j/2}\Gamma S_n/12} \ge \frac{n\gamma^*\Gamma S_n}{192C2^{j/2}} \ge \frac{\gamma^*\Gamma}{192C}2^{-\tilde{j}_n/2}\sqrt{n\log(n)} \ge \frac{\gamma^*\Gamma}{192C}\log(n)$$

since by construction $2^{\tilde{j}_n}\le\frac{n}{\tau^2\log(n)}\le\frac{n}{\log(n)}$. Therefore since $L\le n$ by assumption, for any $A>0$ there exists $c_0>0$ such that whenever $\Gamma\ge c_0\max(L^{1/2}(\gamma^*)^{-1/2},(\gamma^*)^{-1})$:

$$R_{3,1}\le\max\Big(1,\frac{g^2}{m_1^2}\Big)n^{-A}.$$

We now control $R_{3,2}$. With the same argument as before,

$$R_{3,2} \le 36\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\Big(\big(W_2^{\mathfrak{B}_{j\ell}}\big)^2\mathbf{1}_{\{W_2^{\mathfrak{B}_{j\ell}}>\Gamma S_n/12\}}\Big).$$

Proposition 9 tells us that $\mathbb{P}_\theta\Big(\|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| \ge CL\sqrt{\frac{x}{n\gamma^*}} + C\max(\tau 2^{j/2},\sqrt{L}2^{j/2},\tau\sqrt{L})\frac{x}{n\gamma^*}\Big) \le 4\cdot 24^N e^{-x}$. Thus, applying Lemma 8 with $a = \frac{4Cg}{|m_1|\gamma^*}\tau\sqrt{L}2^{j/2}$, $b = \frac{4CLg}{|m_1|\sqrt{\gamma^*}}$, $c = 24^N$, and

$d = \Gamma S_n/12$ [note that $\max(\tau 2^{j/2}, \sqrt{L}2^{j/2}, \tau\sqrt{L}) \leq \tau\sqrt{L}2^{j/2}$], we find that

$$R_{3,2} \leq 36 \cdot 24^N \sum_{j=J_n}^{\tilde{j}_n} \sum_{\ell} \left(\frac{\Gamma^2 S_n^2}{144} + \frac{10C^2L^2g^2}{n\gamma^* m_1^2} + \frac{7 \cdot 4^2 C^2 \tau^2 L 2^j g^2}{2(n\gamma^*)^2 m_1^2} \right) \times \exp\left(- \frac{n\gamma^* \Gamma^2 S_n^2/144}{\frac{8C^2L^2g^2}{m_1^2} + \frac{16C\tau\sqrt{L}2^{j/2}g}{12|m_1|}\Gamma S_n} \right)$$

ie.

$$R_{3,2} \leq 36 \cdot 24^N \max\left(1, \frac{g^2}{m_1^2}\right)\left(\frac{\Gamma^2}{144} + 20C^2L^2n + \frac{14 \cdot 4^2\tau^2 L n^2}{2}\right) \times \exp\left(- \frac{n\gamma^* \Gamma^2 S_n^2/144}{\frac{8C^2L^2g^2}{m_1^2} + \frac{16C\tau\sqrt{L}2^{j/2}g}{12|m_1|}\Gamma S_n} \right)$$

Let us analyse the argument of the exponential in the previous display. Firstly, in the case where $\frac{16C\tau\sqrt{L}2^{j/2}g}{12|m_1|}\Gamma S_n \leq \frac{8C^2L^3g^2}{m_1^2}$,

$$\frac{n\gamma^* \Gamma^2 S_n^2/144}{\frac{8C^2L^2g^2}{m_1^2} + \frac{16C\tau\sqrt{L}2^{j/2}g}{12|m_1|}\Gamma S_n} \geq \frac{n\gamma^*\Gamma^2 S_n^2}{\frac{2304C^2L^2g^2}{m_1^2}} \geq \frac{\gamma^*\Gamma^2}{2304C^2L^2}\log(n)$$

since $S_n = \sqrt{\log(n)/n}\max(1, g/|m_1|)$. Secondly, in the case where $\frac{16C\tau\sqrt{L}2^{j/2}g}{12|m_1|}\Gamma S_n \leq \frac{8C^2L^2g^2}{m_1^2}$, for any $j \leq \tilde{j}_n$

$$\frac{n\gamma^* \Gamma^2 S_n^2/144}{\frac{8C^2L^2g^2}{m_1^2} + \frac{16C\tau\sqrt{L}2^{j/2}g}{12|m_1|}\Gamma S_n} \geq \frac{n\gamma^*\Gamma S_n}{\frac{384C\tau\sqrt{L}2^{j/2}g}{|m_1|}} \geq \frac{\gamma^*\Gamma}{384C\tau\sqrt{L}}2^{-\tilde{j}_n/2}\sqrt{n\log(n)} \geq \frac{\gamma^*\Gamma}{384C\sqrt{L}}\log(n)$$

since by construction $2^{\tilde{j}_n} \leq \frac{n}{\tau^2}\log(n)$. Therefore, for any $A > 0$ there exits a constant $c_0 > 0$ such that whenever $\Gamma \geq c_0 L^{1/2}\max(L^{1/2}(\gamma^*)^{-1/2}, (\gamma^*)^{-1})$

$$R_{3,2} \leq \max\left(1, \frac{g^2}{m_1^2}\right)n^{-A}.$$

We now control $R_{3,3}$. With the same argument as before,

$$R_{3,3} \leq 36 \sum_{j=J_n}^{\tilde{j}_n} \sum_{\ell} \mathbb{E}_\theta\Big(\big(W_3^{\mathfrak{B}_{j\ell}}\big)^2 \mathbf{1}_{\{W_3^{\mathfrak{B}_{j\ell}} > \Gamma S_n/12\}} \mathbf{1}_{\Omega_n} \Big)$$
$$\leq 36 \sum_{j=J_n}^{\tilde{j}_n} \sum_{\ell} \mathbb{E}_\theta\Big(\big(W_3^{\mathfrak{B}_{j\ell}}\big)^2 \mathbf{1}_{\Omega_n} \Big).$$

Proposition 10 in Section A.4.7 tells us that $|\hat{\omega}_\pm - \omega_\pm| \leq \frac{83\max(1, \phi_3|\tilde{\mathcal{I}}|)}{|m_1 m_2|}\max_{j=1,2,3}|\hat{m}_j - m_j|$ on the event Ω_n, hence

$$R_{3,3} \leq \frac{9 \cdot 83^2 \max(1, \phi_3^2 \tilde{\mathcal{I}}^2)}{m_1^2 m_2^2} \mathbb{E}_\theta\Big(\max_{j=1,2,3} |\hat{m}_j - m_j|^2 \Big) \sum_{j=J_n}^{\tilde{j}_n} \sum_{\ell} \|G^{\mathfrak{B}_{j\ell}}\|^2$$

$$\leq \frac{9 \cdot 83^2 \max(1, \phi_3^2 \tilde{\mathcal{I}}^2)}{m_2^2} \mathbb{E}_\theta\Big(\max_{j=1,2,3} |\hat{m}_j - m_j|^2\Big)$$

because $\|G\|_{L^2} = |m_1| \|\psi_2\|_{L^2} = |m_1|$. Furthermore, by Proposition 4, we deduce

$$R_{3,3} \leq \frac{9 \cdot 83^2 \cdot 40 C^2 L^3 \max(1, g^2)}{n\gamma^* m_2^2} + \frac{9 \cdot 83^2 \cdot 64 C^2 \max(\tau, \sqrt{L})^6 \max(1, g^2)}{(n\gamma^*)^2 m_2^2}.$$

In the end for every $A > 0$ there exists $c_0 > 0$ such that whenever the threshold constant satisfies $\Gamma \geq c_0 L^{1/2} \max(L^{1/2}(\gamma^*)^{-1/2}, (\gamma^*)^{-1})$

$$R_3(\theta) \leq 2\max\Big(1, \frac{g^2}{m_1^2}\Big) n^{-A} + \frac{9 \cdot 83^2 \cdot 40 C^2 L^3 \max(1, g^2)}{n\gamma^* m_2^2} + \frac{9 \cdot 83^2 \cdot 64 C^2 \max(\tau, \sqrt{L})^6 \max(1, g^2)}{(n\gamma^*)^2 m_2^2}.$$

By choosing $\beta > 0$ carefully enough, there is a universal constant $B > 0$ such that

$$\sup_{\theta \in \Theta^{s_0, s_1}_{\delta,\epsilon,\zeta}(R) \cap \Sigma_{\gamma^*}(L)} R_3(\theta) \leq \frac{BL^3}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{n\gamma^*} + \frac{B \max(\tau, \sqrt{L})^6}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{(n\gamma^*)^2}.$$

A.4.4. *Control of R_4.* Observe that

$$R_4(\theta) := \mathbb{E}_\theta\Bigg(\sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{S}_n\}} \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| \leq \frac{1}{2}\Gamma \hat{S}_n\}\}} \mathbf{1}_{\Omega_n}\Bigg)$$

$$\leq \mathbb{E}_\theta\Bigg(\sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma \hat{S}_n\}} \mathbf{1}_{\Omega_n}\Bigg)$$

$$\leq \mathbb{E}_\theta\Bigg(\sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{8}\Gamma S_n\}} \mathbf{1}_{\Omega_n}\Bigg)$$

since $\hat{S}_n \geq S_n/4$ on the event Ω_n by Proposition 11 in Section A.4.7. From here, we see that the bounds derived for R_3 adapts mutatis mutandis by letting $\Gamma \mapsto \Gamma/2$. In the end it is found that for $\beta > 0$ chosen carefully enough

$$\sup_{\theta \in \Theta^{s_0, s_1}_{\delta,\epsilon,\zeta}(R) \cap \Sigma_{\gamma^*}(L)} R_4(\theta) \leq \frac{BL^3}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{n\gamma^*} + \frac{B \max(\tau, \sqrt{L})^6}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{(n\gamma^*)^2}.$$

A.4.5. *Control of R_5.* First see that, since $\hat{S}_n \geq S_n/4$ on the event Ω_n by Proposition 11,

$$R_5(\theta) := \mathbb{E}_\theta\Bigg(\sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{S}_n\}} \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma \hat{S}_n\}} \mathbf{1}_{\Omega_n}\Bigg)$$

$$\leq \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\Omega_n}\Big) \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{8}\Gamma S_n\}}.$$

Let $W_j^{\mathfrak{B}_{j\ell}}$ be defined as in Section A.4.3. Then, by Lemma 15 in Section A.4.7,

$$\mathbb{E}_\theta\Big(\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\Omega_n}\Big) \leq 3\mathbb{E}_\theta\Big(\big(W_1^{\mathfrak{B}_{j\ell}}\big)^2\Big) + 3\mathbb{E}_\theta\Big(\big(W_2^{\mathfrak{B}_{j\ell}}\big)^2\Big) + 3\mathbb{E}_\theta\Big(\big(W_3^{\mathfrak{B}_{j\ell}}\big)^2 \mathbf{1}_{\Omega_n}\Big)$$

By computations made in Section A.4.3, for any $A > 0$ we can choose $\alpha > 0$ such that

$$\begin{aligned}\mathbb{E}_\theta\Big(\big(W_1^{\mathfrak{B}_{j\ell}}\big)^2\Big) &\leq \alpha^2 C^2 L \frac{\log(n)}{n\gamma^*} + \mathbb{E}_\theta\Big(\big(W_1^{\mathfrak{B}_{j\ell}}\big)^2 \mathbf{1}_{\{W_1^{\mathfrak{B}_{j\ell}} > \alpha C\sqrt{L\log(n)/(n\gamma^*)}\}}\Big)\\ &\leq \alpha^2 C^2 L \frac{\log(n)}{n\gamma^*} + \max\Big(1, \frac{g^2}{m_1^2}\Big) 2^{-\tilde{\jmath}_n} n^{-A}\\ &\leq \frac{\alpha^2 C^2 L S_n^2}{\gamma^*} + \max\Big(1, \frac{g^2}{m_1^2}\Big) 2^{-\tilde{\jmath}_n} n^{-A}.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}_\theta\Big(\big(W_2^{\mathfrak{B}_{j\ell}}\big)^2\Big) &\leq \alpha^2 C^2 L^2 \frac{g^2\log(n)}{n\gamma^* m_1^2} + \mathbb{E}_\theta\Big(\big(W_2^{\mathfrak{B}_{j\ell}}\big)^2 \mathbf{1}_{\{W_2^{\mathfrak{B}_{j\ell}} > \frac{\alpha C L g}{|m_1|}\sqrt{\log(n)/(n\gamma^*)}\}}\Big)\\ &\leq \alpha^2 C^2 L^2 \frac{g^2\log(n)}{n\gamma^* m_1^2} + \max\Big(1, \frac{g^2}{m_1^2}\Big) 2^{-\tilde{\jmath}_n} n^{-A}\\ &\leq \frac{\alpha^2 C^2 L^2 S_n^2}{\gamma^*} + \max\Big(1, \frac{g^2}{m_1^2}\Big) 2^{-\tilde{\jmath}_n} n^{-A}.\end{aligned}$$

Also, by computations made in Section A.4.3, we know that

$$\begin{aligned}\sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \mathbb{E}_\theta\Big(\big(W_3^{\mathfrak{B}_{j\ell}}\big)^2 \mathbf{1}_{\Omega_n}\Big) \leq \frac{9\cdot 83^2\cdot 40 C^2 L^3 \max(1,g^2)}{36 n\gamma^* m_2^2}&\\ + \frac{9\cdot 83^2\cdot 64 C^2 \max(\tau,\sqrt{L})^6 \max(1,g^2)}{36(n\gamma^*)^2 m_2^2}&.\end{aligned}$$

Consequently,

$$\begin{aligned}R_5(\theta) \leq &\frac{6\alpha^2 C^2 L^2 S_n^2}{\gamma^*}\sum_{j=J_n}^{\tilde{\jmath}_n}\sum_\ell \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{8}\Gamma S_n\}}\\ &+ \frac{27\cdot 83^2\cdot 40C^2L^3\max(1,g^2)}{36n\gamma^* m_2^2} + \frac{27\cdot 83^2\cdot 64 C^2\max(\tau,\sqrt{L})^6\max(1,g^2)}{36(n\gamma^*)^2 m_2^2}\\ &+ 2\max\Big(1,\frac{g^2}{m_1^2}\Big) n^{-A}.\end{aligned}$$

Whenever $\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$, it is the case (recall (26)) that $\sup_{j\geq J_n} 2^{2js_\pm}\sum_k |f_\pm^{\Psi_{jk}}|^2 \leq R^2$. This in particular implies that for all $j \geq J_n$

$$\begin{aligned}R^2 2^{-2js_\pm} &\geq \sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2\\ &\geq \sum_\ell \|f_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{8}\Gamma S_n\}}\\ &\geq \frac{\Gamma^2 S_n^2}{64}\sum_\ell \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{8}\Gamma S_n\}}.\end{aligned}$$

Since there are $2^j/N$ blocks at level j, deduce that

$$\sum_\ell \mathbf{1}_{\{\|f_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{8}\Gamma S_n\}} \leq \min\Big(\frac{2^j}{N}, \frac{64R^2}{\Gamma^2 S_n^2}2^{-2js_\pm}\Big) = \frac{1}{\Gamma^2 S_n^2}\min\Big(\frac{2^j}{N}\Gamma^2 S_n^2, 64R^2 2^{-2js_\pm}\Big)$$

Therefore,

$$
\begin{aligned}
R_5(\theta) \leq{}& \frac{6\alpha^2 C^2 L^2}{\Gamma^2\gamma^*}\sum_{j=J_n}^{\tilde{J}_n}\min\Big(\frac{2^j}{N}\Gamma^2 S_n^2,\, 64R^2 2^{-2js_\pm}\Big)\\
&+\frac{27\cdot 83^2\cdot 40C^2L^3\max(1,g^2)}{36n\gamma^* m_2^2}+\frac{27\cdot 83^2\cdot 64C^2\max(\tau,\sqrt{L})^6\max(1,g^2)}{36(n\gamma^*)^2m_2^2}\\
&+2\max\Big(1,\frac{g^2}{m_1^2}\Big)n^{-A}.
\end{aligned}
$$

Then by inspecting the proof of the bound of $R_2(\theta)$ and by choosing α sufficiently large it follows immediately that there exists a universal constant $B>0$ such that

$$
\begin{aligned}
\sup_{\theta\in\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)\cap\Sigma_{\gamma^*}(L)} R_5(\theta) \leq{}& \frac{BL^3}{\delta^2\epsilon^4\zeta^6}\frac{1}{n\gamma^*}+\frac{B\max(\tau,\sqrt{L})^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2}\\
&+\frac{BL^2}{\Gamma^2\gamma^*}\Bigg(\frac{R^2}{\min(1,s_\pm)}\Big(\frac{\Gamma^2}{R^2\delta^2\epsilon^2\zeta^2 n}\Big)^{2s_\pm/(2s_\pm+1)}+\frac{R^2}{\min(1,s_\pm)}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_\pm}\Bigg).
\end{aligned}
$$

A.4.6. *Control of R_6.* Whenever $\theta\in\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$, it is the case (recall equation (26)) that $\sup_{j\geq J_n}2^{2js_\pm}\sum_k|f_\pm^{\Psi_{jk}}|^2\leq R^2$. Therefore,

$$
\begin{aligned}
R_6(\theta) \coloneqq \mathbb{P}_\theta(\Omega_n)\sum_{j>\tilde{J}_n}\sum_{k=0}^{2^j-1}|f_\pm^{\Psi_{jk}}|^2 \leq R^2\sum_{j>\tilde{J}_n}2^{-2js_\pm} = \frac{L^2}{2^{2s_\pm}-1}2^{-2\tilde{J}_n s_\pm}\\
\leq \frac{R^2}{2^{2s_\pm}-1}\Big(\frac{2\tau^2\log(n)}{n}\Big)^{2s_\pm}
\end{aligned}
$$

because by construction $2^{\tilde{J}_n+1}>\frac{n}{\tau^2\log(n)}$. Hence, there is a universal constant $B>0$ such that

$$
\sup_{\theta\in\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)\cap\Sigma_{\gamma^*}(L)} R_6(\theta)\leq\frac{BR^2}{\min(1,s_\pm)}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_\pm}.
$$

A.4.7. *Auxiliary results.*

PROPOSITION 6. *Let $n\gamma^*\geq 1/99$. Then, there is a universal constant $C>0$ such that for all $\theta\in\Sigma_{\gamma^*}(L)$ and all $x\geq 0$*

$$
\mathbb{P}_\theta\Bigg(\|\hat{\psi}_1^{J_n}-\psi_1^{J_n}\|_{L^2}\geq C\sqrt{\frac{Lx}{n\gamma^*}}+C2^{J_n/2}\frac{x}{n\gamma^*}\Bigg)\leq 24^{2^{J_n}}e^{-x}.
$$

PROOF. The strategy is classical and consists on remarking that $\|\hat{\psi}_1^{J_n}-\psi_1^{J_n}\|_{L^2}=\sup_{u\in U}\langle\hat{\psi}_1^{J_n}-\hat{\psi}_1^{J_n},u\rangle$ where U is the unit ball of the appropriate vector space (which has dimension $2^J+\sum_{j=J}^{J_n-1}2^j=2^{J_n}$). Then, letting $\mathcal{N}$ be a $(1/2)$-net over U and $\pi: U\to\mathcal{N}$ mapping any point $u\in U$ to its closest element in $\mathcal{N}$, we see that

$$
\|\hat{\psi}_1^{J_n}-\psi_1^{J_n}\|_{L^2}=\sup_{u\in U}\langle\hat{\psi}_1^{J_n}-\hat{\psi}_1^{J_n},u\rangle
$$

$$= \sup_{u \in U} \Big(\langle \hat{\psi}_1^{J_n} - \hat{\psi}_1^{J_n}, \pi(u) \rangle + \langle \hat{\psi}_1^{J_n} - \hat{\psi}_1^{J_n}, u - \pi(u) \rangle \Big)$$

$$\leq \max_{u \in \mathcal{N}} \langle \hat{\psi}_1^{J_n} - \hat{\psi}_1^{J_n}, u \rangle + \frac{1}{2} \| \hat{\psi}_1^{J_n} - \psi_1^{J_n} \|_{L^2}$$

and hence

$$\| \hat{\psi}_1^{J_n} - \psi_1^{J_n} \|_{L^2} \leq 2 \max_{u \in \mathcal{N}} \langle \hat{\psi}_1^{J_n} - \hat{\psi}_1^{J_n}, u \rangle.$$

It follows that

$$\mathbb{P}_\theta \Big(\| \hat{\psi}_1^{J_n} - \psi_1^{J_n} \|_{L^2} \geq 2x \Big) \leq |\mathcal{N}| \max_{u \in \mathcal{N}} \mathbb{P}_\theta \Big(\langle \hat{\psi}_1^{J_n} - \hat{\psi}_1^{J_n}, u \rangle \geq x \Big)$$

The conclusion follows by Lemma 7 applied to the function $h(y) = \sum_{k=0}^{2^J-1} u_{Jk} \Phi_{Jk}(y) + \sum_{j=J}^{J_n} \sum_{k=0}^{2^j-1} u_{jk} \psi_{jk}(y)$, because $\mathbb{E}_\theta(h^2) \leq L \|h\|_{L^2}^2 = L$ for every $\theta \in \Sigma_{\gamma^*}(L)$ by Lemma 5, because $\|h\|_\infty \leq c 2^{J_n/2}$ for a universal $c > 0$, by standard localization properties of wavelets [12, Theorem 4.2.10 or Definition 4.2.14] and because $\mathcal{N}$ can be chosen so that $|\mathcal{N}| \leq 24^{2^{J_n}}$ because $\mathcal{N}$ can always be chosen to have cardinality no more than $24^{2^{J_n}}$ (e.g. [12, Theorem 4.3.34]). ☐

PROPOSITION 7. *Let $n\gamma^* \geq 1/99$ and $\|\tilde{\psi}_2\|_\infty \leq \tau$. Then, there is a universal constant $C > 0$ such that for all $\theta \in \Sigma_{\gamma^*}(L)$ and all $x \geq 0$*

$$\mathbb{P}_\theta \left(\| \hat{G}^{J_n} - G^{J_n} \|_{L^2} \geq CL \sqrt{\frac{x}{n\gamma^*}} + C \max(\tau 2^{J_n/2}, \sqrt{L} 2^{J_n/2}, \tau \sqrt{L}) \frac{x}{n\gamma^*} \right) \leq 4 \cdot 24^{2^{J_n}} e^{-x}.$$

PROOF. We remark that $\hat{G}^{\Phi_{Jk}} = \mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \Phi_{Jk}) - \mathbb{P}_n^{(1)}(\tilde{\psi}_2) \mathbb{P}_n^{(1)}(\Phi_{Jk})$; similarly for $\hat{G}^{\Psi_{jk}}$. Recall that $\|\tilde{\psi}_2\|_\infty \leq \tau$ by assumption. Hence, $\|\hat{G}^{J_n}\|_{L^2} \leq c\tau 2^{J_n/2}$ for a universal constant $c > 0$. Similarly $\|G^{J_n}\|_{L^2} \leq c\tau 2^{J_n/2}$. Hence with probability $1 \geq 1 - e^{-x}$, whenever $x > n\gamma^*$

$$\| \hat{G}^{J_n} - G^{J_n} \|_{L^2} \leq 2c\tau 2^{J_n/2} \leq CL^{3/2} \sqrt{\frac{x}{n\gamma^*}}$$

provided $C > 2c$. We now consider the case where $0 \leq x \leq n\gamma^*$. We decompose

$$\begin{aligned} \hat{G}^{J_n} - G^{J_n} = & \sum_{k=0}^{2^J-1} \Big(\mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \Phi_{Jk}) - \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \Phi_{Jk}) \Big) \Phi_{Jk} \\ & + \sum_{j=J}^{J_n} \sum_{k=0}^{2^j-1} \Big(\mathbb{P}_n^{(2)}(\tilde{\psi}_2 \otimes \Psi_{jk}) - \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \Psi_{jk}) \Big) \Psi_{jk} \\ & - \mathbb{E}_\theta(\tilde{\psi}_2) \Big(\hat{\psi}_1^{J_n} - \psi_1^{J_n} \Big) - \psi_1^{J_n} \Big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2) \Big) \\ & - \Big(\mathbb{P}_n^{(1)}(\tilde{\psi}_2) - \mathbb{E}_\theta(\tilde{\psi}_2) \Big) \Big(\hat{\psi}_1^{J_n} - \psi_1^{J_n} \Big). \end{aligned}$$

But $\|\psi_1^{J_n}\|_{L^2} \leq \|\psi_1\|_{L^2} \leq \max(\|f_0\|_{L^2}, \|f_1\|_{L^2})$ and $\|f_j\|_{L^2}^2 = \int_0^1 f_j^2 \leq \|f_j\|_\infty \int_0^1 f_j \leq L$ whenever $\theta \in \Sigma_{\gamma^*}(L)$. Thus $\|\psi_1^{J_n}\|_{L^2} \leq \sqrt{L}$. Similarly by Cauchy-Schwarz' $|\mathbb{E}_\theta(\tilde{\psi}_2)| \leq \mathbb{E}_\theta(\tilde{\psi}_2^2)^{1/2} \leq \|\psi_1\|_\infty^{1/2} \|\tilde{\psi}_2\|_{L^2} \leq \sqrt{L}$. Therefore, letting $v^{J_n} := \sum_{k=0}^{2^J-1} \mathbb{E}_\theta(\tilde{\psi}_2 \otimes \Phi_{Jk}) \Phi_{Jk} +$

$\sum_{j=J}^{J_n}\sum_{k=0}^{2^j-1}\mathbb{E}_\theta(\tilde\psi_2\otimes\Psi_{jk})\Psi_{jk}$ and its empirical counterpart $\hat v^{J_n}$ defined similarly:

$$\|\hat G^{J_n}-G^{J_n}\|_{L^2}\le\|\hat v^{J_n}-v^{J_n}\|_{L^2}+\sqrt{L}\|\hat\psi_1^{J_n}-\psi_1^{J_n}\|_{L^2}+\sqrt{L}\Big|\mathbb{P}_n^{(1)}(\tilde\psi_2)-\mathbb{E}_\theta(\tilde\psi_2)\Big| \\ +\Big|\mathbb{P}_n^{(1)}(\tilde\psi_2)-\mathbb{E}_\theta(\tilde\psi_2)\Big|\|\hat\psi_1^{J_n}-\psi_1^{J_n}\|_{L^2}.$$

Using the same ε-net argument as in the proof of Proposition 6, we find that

$$\mathbb{P}_\theta\Big(\|\hat v^{J_n}-v^{J_n}\|_{L^2}\ge CL\sqrt{\frac{x}{n\gamma^*}}+C\tau 2^{J/2}\frac{x}{n\gamma^*}\Big) \\ \le 24^{2^{J_n}}\sup_{u\in U}\mathbb{P}_\theta\Big(\langle\hat v^{J_n}-v^{J_n},u\rangle\ge CL\sqrt{\frac{x}{n\gamma^*}}+C\tau 2^{J/2}\frac{x}{n\gamma^*}\Big)\le 24^{2^{J_n}}e^{-x}$$

where the last inequality follows from Lemma 7 applied to the function $h(y_1,y_2)=\sum_{k=0}^{2^J-1}u_{Jk}\tilde\psi_2(y_1)\Phi_{Jk}(y_2)+\sum_{j=J}^{J_n}\sum_{k=0}^{2^j-1}u_{jk}\tilde\psi_2(y_1)\Psi_{jk}(y_1)$ which satisfies $\mathbb{E}_\theta(h^2)\le L^2\|h\|_{L^2}^2=L^2$ for every $\theta\in\Sigma_{\gamma^*}(L)$ by Lemma 5, and $\|h\|_\infty\le c\|\tilde\psi_2\|_\infty 2^{J/2}\le c\tau 2^{J/2}$ by standard localization properties of wavelets [12, Theorem 4.2.10 or Definition 4.2.14]. Also by Lemma 7 applies to $h=\tilde\psi_2$,

$$\mathbb{P}_\theta\left(\Big|\mathbb{P}_n^{(1)}(\tilde\psi_2)-\mathbb{E}_\theta(\tilde\psi_2)\Big|\ge C\sqrt{\frac{Lx}{n\gamma^*}}+C\tau\frac{x}{n\gamma^*}\right)\le e^{-x}$$

and using Proposition 6

$$\mathbb{P}_\theta\left(\|\hat\psi_1^{J_n}-\psi_1^{J_n}\|\ge C\sqrt{\frac{Lx}{n\gamma^*}}+C2^{J/2}\frac{x}{n\gamma^*}\right)\le 24^{2^{J_n}}e^{-x}.$$

Therefore with probability at least $1-(2\cdot 24^{2^{J_n}}+1)e^{-x}$ under $\mathbb{P}_\theta$

$$\|\hat G^{J_n}-G^{J_n}\|_{L^2}\le C\left(\sqrt{\frac{L^2x}{n\gamma^*}}+\tau 2^{J_n/2}\frac{x}{n\gamma^*}\right)+C\sqrt{L}\left(\sqrt{\frac{Lx}{n\gamma^*}}+2^{J_n/2}\frac{x}{n\gamma^*}\right) \\ +C\sqrt{L}\left(\sqrt{\frac{Lx}{n\gamma^*}}+\tau\frac{x}{n\gamma^*}\right)+C^2\left(\sqrt{\frac{Lx}{n\gamma^*}}+2^{J_n/2}\frac{x}{n\gamma^*}\right)\left(\sqrt{\frac{Lx}{n\gamma^*}}+\tau\frac{x}{n\gamma^*}\right) \\ \le 3CL\sqrt{\frac{x}{n\gamma^*}}+C\Big(\tau 2^{J_n/2}+\sqrt{L}2^{J_n/2}+\tau\sqrt{L}+CL\Big)\frac{x}{n\gamma^*} \\ +C^2\Big(\tau\sqrt{L}+2^{J_n/2}\sqrt{L}\Big)\frac{x^{3/2}}{(n\gamma^*)^{3/2}}+C^2\tau 2^{J_n/2}\frac{x^2}{(n\gamma^*)^2}.$$

The conclusion follows since $x\le n\gamma^*$ which implies that the last two terms are bounded by the second term, and the $Lx/(n\gamma^*)$ part of second term is bounded by the first term. □

PROPOSITION 8. *Let $n\gamma^*\ge 1/99$. Then, there is a universal constant $C>0$ such that for all $\theta\in\Sigma_{\gamma^*}(L)$, all $j\ge J_n$, all ℓ, and all $x\ge 0$,*

$$\mathbb{P}_\theta\left(\|\hat\psi_1^{\mathfrak{B}_{j\ell}}-\psi_1^{\mathfrak{B}_{j\ell}}\|\ge C\sqrt{\frac{Lx}{n\gamma^*}}+C2^{j/2}\frac{x}{n\gamma^*}\right)\le 24^N e^{-x}.$$

PROOF. The proof is identical to Proposition 6. (Note the vector $\psi_1^{\mathfrak{B}_{j\ell}}$ is in $\mathbb{R}^N$, where ψ_1^{Φ} was in $\mathbb{R}^{2^{J_n}}$.) □

PROPOSITION 9. *Let $n\gamma^* \geq 1/99$. Then, there is a universal constant $C > 0$ such that for all $\theta \in \Sigma_{\gamma^*}(L)$, all $j \geq J_n$, all ℓ, and all $x \geq 0$*

$$\mathbb{P}_\theta\left(\|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| \geq CL\sqrt{\frac{x}{n\gamma^*}} + C\max(\tau 2^{j/2}, \sqrt{L}2^{j/2}, \tau\sqrt{L})\frac{x}{n\gamma^*}\right) \leq 4 \cdot 24^N e^{-x}.$$

PROOF. The proof is identical to Proposition 7. □

LEMMA 15. *On the event Ω_n, for all $j \geq J_n$ and all ℓ:*

$$\|\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}}\| \leq \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| + \frac{4g\|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\|}{|m_1|} + \frac{|\hat{\omega}_\pm - \omega_\pm|\|G^{\mathfrak{B}_{j\ell}}\|}{2},$$

and similarly for $\|\hat{f}_\pm^{J_n} - f_\pm^{J_n}\|_{L^2}$.

PROOF. Trivially,

$$\hat{f}_\pm^{\mathfrak{B}_{j\ell}} - f_\pm^{\mathfrak{B}_{j\ell}} = \hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}} + \frac{\hat{\omega}_\pm}{2}\left(\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\right) + \frac{\hat{\omega}_\pm - \omega_\pm}{2}G^{\mathfrak{B}_{j\ell}}.$$

The conclusion follows since on Ω_n, Proposition 11 implies that $\hat{g} \leq 2g$ and $|\hat{m}_1| \geq |m_1|/2 > 0$. (Recall also that $|\phi_1| \leq 1$.) □

PROPOSITION 10. *On the event Ω_n*

$$|\hat{\omega}_\pm - \omega_\pm| \leq \frac{83\max(1, \phi_3|\tilde{\mathcal{I}}|)}{|m_1 m_2|}\max_{j=1,2,3}|\hat{m}_j - m_j|.$$

PROOF. On Ω_n we have $\hat{g} \leq 2g$ by Proposition 11 to follow, and note that $|\hat{m}_1| \geq |m_1|/2 > 0$. Consequently, by straightforward computations, using Lemmas 12 and 16,

$$\begin{aligned}
|\hat{\omega}_\pm - \omega_\pm| &= \Big|\frac{1}{m_1}(\hat{g} - g)(1 \mp \hat{\phi}_1) + \frac{g}{m_1}(1 \mp \hat{\phi}_1 - (1 \mp \tilde{s}\phi_1)) + \hat{g}(1 \mp \hat{\phi}_1)(\frac{1}{\hat{m}_1} - \frac{1}{m_1})\Big| \\
&\leq \frac{2|\hat{g} - g|}{|m_1|} + \frac{g|\hat{\phi}_1 - \tilde{s}\phi_1|}{|m_1|} + \frac{8g|\hat{m}_1 - m_1|}{m_1^2} \\
&\leq \left(\frac{22\max(1, \phi_3|\tilde{\mathcal{I}}|)}{|m_1 m_2|} + \frac{53\max(1, \phi_3|\tilde{\mathcal{I}}|)g}{|m_1|\phi_2^2\phi_3^3|\tilde{\mathcal{I}}|^3} + \frac{8g}{m_1^2}\right)\max_{j=1,2,3}|\hat{m}_j - m_j| \\
&\leq \frac{83\max(1, \phi_3|\tilde{\mathcal{I}}|)}{|m_1 m_2|}\max_{j=1,2,3}|\hat{m}_j - m_j|
\end{aligned}$$

because $m_2 = \frac{1}{4}(1 - \phi_1^2)\phi_2^2\phi_3^2\tilde{\mathcal{I}}^2$, because $g = \phi_3|\tilde{\mathcal{I}}|$, and because $m_2 = m_1\phi_2 \leq |m_1|$. □

PROPOSITION 11. *On the event Ω_n, we have $|\frac{\hat{g}}{g} - 1| \leq \frac{1}{2}$. Consequently, $\frac{1}{4}S_n \leq \hat{S}_n \leq 4S_n$ and $|\hat{\omega}_\pm| \leq 8|g/m_1|$ on Ω_n.*

PROOF. It suffices to remark that

$$\frac{gm_2}{44\max(1, g)} \leq \frac{gm_2}{20\max(|\phi_1|, (1 - \phi_1^2)\phi_3|\tilde{\mathcal{I}}|)},$$

since $-1 \le \phi_1 \le 1$, so that Lemma 16 applies. Replacing $\max_j|\hat{m}_j - m_j|$ by its bound $gm_2/44\max(1,g)$ on the event Ω_n yields the result for $\hat{g}$. For S_n, recalling the definitions $S_n = \sqrt{(\log n)/n}\max(1, g/|m_1|)$, $\hat{S}_n = \sqrt{(\log n)/n}\max(1, \hat{g}/|\hat{m}_1|)\mathbb{1}\{\hat{m}_1 \neq 0\}$ and inserting the bounds $g/2 \le \hat{g} \le 2g$, $|m_1|/2 \le \hat{m}_1 \le 2|m_1|$ yields the bounds for $\hat{S}_n$. ☐

LEMMA 16. *Suppose*

$$\max_{j=1,2}\left|\frac{\hat{m}_j}{m_j} - 1\right| \le \frac{1}{2}, \quad \text{and,} \quad \max_{j=1,2,3}|\hat{m}_j - m_j| \le \frac{m_2 g}{20\max(|\phi_1|, (1-\phi_1^2)g)}$$

Then,

$$|\hat{g} - g| \le \frac{22\max(1,g)}{m_2}\max_{j=1,2,3}|\hat{m}_j - m_j|.$$

Recall that $g = \phi_3|\tilde{\mathcal{I}}|$ and $m_2 = \phi_2 r(\phi)\tilde{\mathcal{I}}^2$, so that the conditions of Lemma 16 match those of Lemma 12.

PROOF. We let $\Delta_1 = \hat{m}_1 - m_1$, $\Delta_2 = (\hat{m}_2)_+ - m_2$, $\Delta_3 = \hat{m}_3 - m_3$, $\hat{v} := 4\hat{m}_1^2(\hat{m}_2)_+ + \hat{m}_3^2$, $v := 4m_1^2 m_2 + m_3^2$, and $h := \hat{v} - v$. Then, since $\hat{m}_2 \ge m_2/2 > 0$ under the assumption of the lemma

$$\begin{aligned}
\hat{g} - g &= \frac{\sqrt{v+h}}{m_2 + \Delta_2} - \frac{\sqrt{v}}{m_2} \\
&= \frac{\sqrt{v+h} - \sqrt{v}}{m_2 + \Delta_2} - \frac{\Delta_2\sqrt{v}}{m_2(m_2+\Delta_2)} \\
&= \frac{h}{(\sqrt{v+h} + \sqrt{v})(m_2+\Delta_2)} - \frac{\Delta_2\sqrt{v}}{m_2(m_2+\Delta_2)}.
\end{aligned}$$

Hence it must be that

$$|\hat{g} - g| \le \frac{2|h|}{m_2\sqrt{v}} + \frac{2\sqrt{v}}{m_2^2}|\Delta_2|.$$

Lemma 13, together with the fact that $|\phi_1| \le 1$, tells us that

$$|h| \le 10\max(1, \phi_3|\tilde{\mathcal{I}}|)|r(\phi)\phi_2\phi_3\tilde{\mathcal{I}}^3|\max_{j=1,2,3}|\Delta_j|$$

and $\sqrt{v} = |\tilde{\mathcal{I}}|^3 r(\phi)\phi_2\phi_3 = |\tilde{\mathcal{I}}|m_2\phi_3 \le m_2\max(1, \phi_3|\tilde{\mathcal{I}}|)$, thus

$$|\hat{g} - g| \le \frac{20\max(1, \phi_3|\tilde{\mathcal{I}}|)}{m_2}\max_{j=1,2,3}|\Delta_j| + \frac{2\max(1, \phi_3|\tilde{\mathcal{I}}|)}{m_2}|\Delta_2|$$

concluding the proof. ☐

A.5. Proof of Theorem 5.

A.5.1. *Definitions and rationale.* To avoid issues with the non-identifiability, we once again define $p_\pm$ and $f_\pm$ as in Lemma 1. The starting point of the proof is to remark that $f_\pm$ can be rewritten as

$$f_\pm = \left[\frac{2\psi_1}{1 \pm \tilde{s}\phi_1}\right] + \left[-\left(\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\psi_1 \mp \frac{g(1 \mp \tilde{s}\phi_1)}{2m_1}G\right)\right].$$

Then each of the two functions in bracket;s in the previous display is estimated separately using block-thresholded wavelets estimators. The population mother coefficients are defined as

$$\alpha_{\pm}^{\Psi_{jk}} := \frac{2\psi_1^{\Psi_{jk}}}{1 \pm \tilde{s}\phi_1}, \qquad \beta_{\pm}^{\Psi_{jk}} := -\left(\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\psi_1^{\Psi_{jk}} \mp \frac{g(1 \mp \tilde{s}\phi_1)}{2m_1} G^{\Psi_{jk}}\right)$$

and the corresponding empirical versions are

$$\hat{\alpha}_{\pm}^{\Psi_{jk}} := \frac{2\hat{\psi}_1^{\Psi_{jk}}}{1 \pm \hat{\phi}_1}\mathbf{1}_{\{\hat{\phi}_1 \neq \mp 1\}}, \qquad \hat{\beta}_{\pm}^{\Psi_{jk}} := -\left(\frac{1 \mp \hat{\phi}_1}{1 \pm \hat{\phi}_1}\mathbf{1}_{\{\hat{\phi}_1 \neq \mp 1\}}\hat{\psi}_1^{\Psi_{jk}} \mp \frac{\hat{g}(1 \mp \hat{\phi}_1)}{2\hat{m}_1}\mathbf{1}_{\{\hat{m}_1 \neq 0\}}\hat{G}^{\Psi_{jk}}\right).$$

Then, the untruncated estimators can be rewritten as (here $\hat{f}_{\pm}^{\Phi_{Jk}}$ are the father coefficients that were defined in the beginning of Section A.4)

$$\begin{aligned}
\hat{f}_{\pm}^{R} :=& \sum_{k=0}^{2^{J_n}-1} \hat{f}_{\pm}^{\Phi_{Jk}} \Phi_{J_n k} + \sum_{j=J}^{J_n-1}\sum_{k=0}^{2^j-1} \hat{f}_{\pm}^{\Psi_{jk}} \Psi_{jk} \\
&+ \sum_{j=J_n}^{\tilde{\jmath}_n} \sum_{\ell=0}^{2^j/N-1} \left(\sum_{k \in \mathfrak{B}_{j\ell}} \hat{\alpha}_{\pm}^{\Psi_{jk}} \Psi_{jk}\right) \mathbf{1}_{\{\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| > \Gamma\sqrt{\log(n)/n}\}} \\
&+ \sum_{j=J_n}^{\tilde{\jmath}_n} \sum_{\ell=0}^{2^j/N-1} \left(\sum_{k \in \mathfrak{B}_{j\ell}} \hat{\beta}_{\pm}^{\Psi_{jk}} \Psi_{jk}\right) \mathbf{1}_{\{\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma\hat{T}_n\}}
\end{aligned}$$

while the truncated versions are

$$\check{f}_{\pm}^{R} := \max\left(0, \min\left(\check{T}, \hat{f}_{\pm}^{R}\right)\right).$$

A.5.2. *Decomposition of the error.* We define auxiliary events

$$\Xi_n^{(1)} := \left\{\forall j = J_n, \dots, \tilde{\jmath}_n,\ \forall \ell,\ \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| \le c_0 \Gamma\sqrt{\log(n)/n}\right\},$$

and

$$\Xi_n^{(2)} := \left\{\forall j = J_n, \dots, \tilde{\jmath}_n,\ \forall \ell,\ \|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| \le c_1 \Gamma\sqrt{\log(n)/n}\right\}$$

. We let Ξ_n denote the intersection of both of these events. Then by the same argument used in Section A.4

$$\mathbb{E}_\theta\left(\|\check{f}_{\pm}^{R} - f_{\pm}\|_{L^2}^2\right) \le \check{T}^2\left(\mathbb{P}_\theta(\Omega_n^c) + \mathbb{P}_\theta(\Xi_n^c)\right) + \mathbb{E}_\theta\left(\|\hat{f}_{\pm}^{R} - f_{\pm}\|_{L^2}^2 \mathbf{1}_{\Omega_n \cap \Xi_n}\right).$$

The probability of the event Ω_n^c is bounded in Proposition 5, while the probability of Ξ_n^c is bounded in Lemma 17 (to follow). We bound the remaining term by decomposing it into several terms. For this matter, we introduce the events

$$E_{j\ell} := \left\{\max_{j=1,2,3} |\hat{m}_j - m_j| \|G^{\mathfrak{B}_{j\ell}}\| \le c_2 |m_1 m_2| \Gamma T_n / \max(1, g)\right\}$$

and we decompose

$$\begin{aligned}
\mathbb{E}_\theta\left(\|\hat{f}_{\pm}^{R} - f_{\pm}\|_{L^2}^2 \mathbf{1}_{\Omega_n \cap \Xi_n}\right) &= \mathbb{E}_\theta\left(\|\hat{f}_{\pm}^{J_n} - f_{\pm}^{J_n}\|_{L^2}^2 \mathbf{1}_{\Omega_n \cap \Xi_n}\right) \\
&+ \sum_{j=J_n}^{\tilde{\jmath}_n} \sum_{\ell} \mathbb{E}_\theta\Big(\left\|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\right\|^2 \\
&\qquad \times \mathbf{1}_{\Omega_n \cap \Xi_n} \mathbf{1}_{n\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|^2 > \Gamma^2 \log(n)} \mathbf{1}_{\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma\hat{T}_n} \mathbf{1}_{E_{j\ell}^c}\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\Big(\|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}}+\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}-\alpha_\pm^{\mathfrak{B}_{j\ell}}-\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2\\
&\qquad\qquad\times\mathbf{1}_{\Omega_n\cap\Xi_n}\mathbf{1}_{n\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|^2>\Gamma^2\log(n)}\mathbf{1}_{\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma\hat{T}_n}\mathbf{1}_{E_{j\ell}}\Big)\\
&+\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\Big(\|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}}-\alpha_\pm^{\mathfrak{B}_{j\ell}}-\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2\\
&\qquad\qquad\times\mathbf{1}_{\Omega_n\cap\Xi_n}\mathbf{1}_{n\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|^2>\Gamma^2\log(n)}\mathbf{1}_{\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}\|\le\Gamma\hat{T}_n}\mathbf{1}_{E_{j\ell}}\mathbf{1}_{\|\psi_1^{\mathfrak{B}_{j\ell}}\|>\frac{g(1\pm\tilde{s}\phi_1)}{|m_1|}\|G^{\mathfrak{B}_{j\ell}}\|}\Big)\\
&+\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\Big(\|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}}-\alpha_\pm^{\mathfrak{B}_{j\ell}}-\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2\\
&\qquad\qquad\times\mathbf{1}_{\Omega_n\cap\Xi_n}\mathbf{1}_{n\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|^2>\Gamma^2\log(n)}\mathbf{1}_{\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}\|\le\Gamma\hat{T}_n}\mathbf{1}_{E_{j\ell}}\mathbf{1}_{\|\psi_1^{\mathfrak{B}_{j\ell}}\|\le\frac{g(1\pm\tilde{s}\phi_1)}{|m_1|}\|G^{\mathfrak{B}_{j\ell}}\|}\Big)\\
&+\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\big(\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}-\alpha_\pm^{\mathfrak{B}_{j\ell}}-\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\Omega_n\cap\Xi_n}\mathbf{1}_{n\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|^2\le\Gamma^2\log(n)}\mathbf{1}_{\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}\|>\Gamma\hat{T}_n}\mathbf{1}_{E_{j\ell}}\big)\\
&+\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\mathbb{E}_\theta\big(\|\alpha_\pm^{\mathfrak{B}_{j\ell}}+\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\Omega_n\cap\Xi_n}\mathbf{1}_{n\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|^2\le\Gamma^2\log(n)}\mathbf{1}_{\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}}\|\le\Gamma\hat{T}_n}\mathbf{1}_{E_{j\ell}}\big)\\
&+\sum_{j>\tilde{j}_n}\sum_{k}|f_\pm^{\Psi_{jk}}|^2\mathbb{P}_\theta(\Omega_n\cap\Xi_n)
\end{aligned}
$$

where we have used the same convention as in Section A.4 to define $\hat{f}_\pm^{J_n}$ and $f_\pm^{J_n}$. We call $R_1(\theta),\dots,R_8(\theta)$, respectively, each of the terms of the previous right hand side. In the next subsections, after stating a couple of preliminary results, we prove the following bounds, uniformly over $\theta\in\Theta_{\delta,\epsilon,\zeta}^{s_0,\tilde{s}_1}(R)\cap\Sigma_{\gamma^*}(L)$ and for a universal constant $B>0$:

$$
\begin{aligned}
R_1(\theta)&\le\frac{BL^2}{\delta^2\epsilon^2\zeta^2}\frac{\log(n)}{n\gamma^*}+\frac{BL^3}{\delta^2\epsilon^4\zeta^4}\frac{1}{n\gamma^*}+\frac{B\max(\tau,L)^6}{\delta^2\epsilon^4\zeta^4}\frac{1}{(n\gamma^*)^2}.\\
R_2(\theta)&\le\frac{B}{\delta^2\epsilon^4\zeta^4}\Big(\frac{L^3}{n\gamma^*}+\frac{\max(\tau,L)^6}{(n\gamma^*)^2}\Big).\\
R_3(\theta)&\le\frac{BR^2}{\min(1,s_\mp)}\frac{1}{\delta^2}\Big(\frac{\Gamma^2}{R^2n\epsilon^2\zeta^2}\Big)^{2s_\mp/(2s_\mp+1)}.\\
R_4(\theta)&\le\frac{BR^2}{\min(1,s_\pm)}\Big(\frac{\Gamma^2}{nR^2\delta^2}\Big)^{2s_\pm/(2s_\pm+1)}+\frac{BR^2}{\min(1,s_\mp)}\frac{1}{\delta^2}\Big(\frac{\Gamma^2}{R^2n\epsilon^2\zeta^2}\Big)^{2s_\mp/(2s_\mp+1)}.\\
R_5(\theta)&\le\frac{B}{\delta^2\epsilon^4\zeta^4}\Big(\frac{L^3}{n\gamma^*}+\frac{\max(\tau,L)^6}{(n\gamma^*)^2}\Big)+\frac{BR^2}{\min(1,s_\pm)}\Big(\frac{\Gamma^2}{nR^2\delta^2}\Big)^{2s_\pm/(2s_\pm+1)}\\
&\quad+\frac{BR^2}{\min(1,s_\mp)}\frac{1}{\delta^2}\Big(\frac{\Gamma^2}{R^2n\epsilon^2\zeta^2}\Big)^{2s_\mp/(2s_\mp+1)}\\
R_6(\theta)&\le\frac{BR^2}{\min(1,s_\mp)}\frac{1}{\delta^2}\Big(\frac{\Gamma^2}{R^2n\epsilon^2\zeta^2}\Big)^{2s_\mp/(2s_\mp+1)}.
\end{aligned}
$$

$$R_7(\theta) \leq \frac{BR^2}{\min(1, s_\pm)}\Big(\frac{\Gamma^2}{nR^2\delta^2}\Big)^{2s_\pm/(2s_\pm+1)} + \frac{R^2}{\min(1, s_\mp)}\frac{1}{\delta^2}\Big(\frac{\Gamma^2}{BR^2 n\epsilon^2\zeta^2}\Big)^{2s_\mp/(2s_\mp+1)}.$$

$$R_8(\theta) \leq \frac{BR^2}{\min(1, s_\pm)}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_\pm}.$$

A.5.3. *Preliminary computations.*

LEMMA 17. *For all $A > 0$ and for all choice of $c_0, c_1 > 0$ there exists a constant $\beta_0 > 0$ such that if $\Gamma \geq \beta\max(\frac{L}{\sqrt{\gamma^*}}, \frac{\sqrt{L}}{\tau\gamma^*})$ with $\beta \geq \beta_0$ then*

$$\mathbb{P}_\theta(\Xi_n^c) \leq n^{-A}.$$

PROOF. By a union bound,

$$\begin{aligned}
\mathbb{P}_\theta\big((\Xi_n^{(1)})^c\big) &\leq \sum_{j=J_n}^{\tilde{J}_n}\sum_\ell \mathbb{P}_\theta\Big(\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| > c_0\Gamma\sqrt{\log(n)/n}\Big)\\
&\leq \frac{2^{\tilde{J}_n+1}}{N}\max_{j\leq\tilde{J}_n}\max_\ell \mathbb{P}_\theta\Big(\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| > c_0\Gamma\sqrt{\log(n)/n}\Big)\\
&\leq n\max_{j\leq\tilde{J}_n}\max_\ell \mathbb{P}_\theta\Big(\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| > c_0\Gamma\sqrt{\log(n)/n}\Big).
\end{aligned}$$

Then choose $x = B\log(n)$ for some $B > 0$ to be chosen accordingly. Observe that for all $j \leq \tilde{J}_n$ (recall $L \geq 1$)

$$\begin{aligned}
C\sqrt{\frac{Lx}{n\gamma^*}} + C2^{j/2}\frac{x}{n\gamma^*} &\leq \frac{C\sqrt{BL}}{\sqrt{\gamma^*}}\cdot\sqrt{\frac{\log(n)}{n}} + C\sqrt{\frac{n}{\log(n)\tau^2}}\frac{B\log(n)}{n\gamma^*}\\
&\leq \frac{C\sqrt{B}+CB}{\beta}\Gamma\sqrt{\log(n)/n}.
\end{aligned}$$

Hence by choosing $c_0 = (C\sqrt{B}+CB)/\beta$ we deduce from the Proposition 8 that

$$\mathbb{P}_\theta\big((\Xi_n^{(1)})^c\big) \leq 24^N n^{-B+1}.$$

The probability of $\Xi_n^{(2)}$ is bounded similarly, remarking that for $x = B\log(n)/n$ we have for all $j \leq \tilde{J}_n$

$$\begin{aligned}
&CL\sqrt{\frac{x}{n\gamma^*}} + C\max(\tau 2^{j/2}, \sqrt{L}2^{j/2}, \tau\sqrt{L})\frac{x}{n\gamma^*}\\
&\qquad\leq \frac{CL\sqrt{B}}{\sqrt{\gamma^*}}\sqrt{\frac{\log(n)}{n}} + \frac{CB}{\gamma^*}\max\left(\sqrt{\frac{n}{\log(n)}}, \frac{\sqrt{L}}{\tau}\sqrt{\frac{\log(n)}{n}}, \tau\sqrt{L}\right)\frac{\log(n)}{n}\\
&\qquad\leq \frac{CL\sqrt{B}}{\sqrt{\gamma^*}}\sqrt{\frac{\log(n)}{n}} + \frac{CB\sqrt{L}}{\gamma^*\tau}\sqrt{\frac{\log(n)}{n}}\\
&\qquad\leq \frac{C\sqrt{B}+CB}{\beta}\Gamma\sqrt{\log(n)/n},
\end{aligned}$$

where the third line is true because by assumption $1 \leq 2^{J_n} \leq 2^{\tilde{J}_n} \leq \frac{n}{\log(n)\tau^2}$ and hence $\tau \leq \sqrt{n/\log(n)}$ necessarily. We then deduce from Proposition 9 that

$$\mathbb{P}_\theta\big((\Xi_n^{(1)})^c\big) \leq 4\cdot 24^N n^{-B+1}$$

which concludes the proof by taking B sufficiently large. $\square$

LEMMA 18. *On the event* Ω_n

$$\frac{1}{2} \le \frac{1 \pm \tilde{s}\phi_1}{1 \pm \hat{\phi}_1} \le 2, \qquad \text{and,} \qquad \frac{1}{2} \le \frac{1 \mp \tilde{s}\phi_1}{1 \mp \hat{\phi}_1} \le 2.$$

PROOF. Observe that

$$\frac{1 \pm \tilde{s}\phi_1}{1 \pm \hat{\phi}_1} = \frac{1}{1 + \frac{\hat{\phi}_1 - \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}}$$

But on the event Ω_n, by Lemma 12

$$\begin{aligned} |\hat{\phi}_1 - \tilde{s}\phi_1| &\le \frac{53\max(1,g)}{gm_2} \cdot \frac{1-\phi_1^2}{4} \cdot \max_{j=1,2,3} |\hat{m}_j - m_j| \\ &\le \frac{53}{4 \cdot 44}(1 \pm \tilde{s}\phi_1)(1 \mp \tilde{s}\phi_1) \\ &\le \frac{1 \pm \tilde{s}\phi_1}{2} \end{aligned}$$

which proves the first claim. The second claim is proven similarly. $\square$

LEMMA 19. *On the event* Ω_n *we have* $\hat{m}_1 \neq 0$ *and* $\hat{\phi}_1^2 \neq 1$.

PROOF. The fact that $\hat{m}_1 \neq 0$ follows immediately from the definition of Ω_n. The fact that $\hat{\phi}_1^2 \neq 1$ follows from Lemma 18 (either one of the two inequalities would not hold if $\hat{\phi}_1^2 = 1$). $\square$

The next Proposition controls the empirical threshold $\hat{T}_n$ in term of its population version defined as

$$T_n := \sqrt{\frac{\log(n)}{n}} \max\Big(1, \frac{g}{|m_1|}, \frac{1}{1-\phi_1^2}\Big).$$

LEMMA 20. *On the event* Ω_n, $\frac{1}{4}T_n \le \hat{T}_n \le 4T_n$.

PROOF. Notice that $T_n = \max\big(S_n, \frac{\sqrt{\log(n)/n}}{1-\phi_1^2}\big)$. Thus, in view of Proposition 11 it is enough to show that $\frac{1-\phi_1^2}{4} \le 1 - \hat{\phi}_1^2 \le 4(1-\phi_1^2)$. But,

$$1-\hat{\phi}_1^2 = (1 \pm \hat{\phi}_1)(1 \mp \hat{\phi}_1) = \frac{1 \pm \hat{\phi}_1}{1 \pm s\tilde{\phi}_1}\frac{1 \mp \hat{\phi}_1}{1 \mp \tilde{s}\phi_1}(1 \mp \tilde{s}\phi_1)(1 \pm \tilde{s}\phi_1) = \frac{1 \pm \hat{\phi}_1}{1 \pm s\tilde{\phi}_1}\frac{1 \mp \hat{\phi}_1}{1 \mp \tilde{s}\phi_1}(1-\phi_1^2).$$

Thus the conclusion follows from Lemma 18. $\square$

LEMMA 21. *It is possible to choose* c_0, c_1, c_2 *such that on the event* $E_{j\ell} \cap \Xi_n \cap \Omega_n$*:*

1. $\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma\hat{T}_n \implies \|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\| > \frac{1}{32}\Gamma T_n$;
2. $\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| \le \Gamma\hat{T}_n \implies \|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\| \le 32\Gamma T_n$;
3. $\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| > \Gamma\sqrt{\log(n)/n} \implies \|\psi_1^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma\sqrt{\log(n)/n}$;
4. $\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| \le \Gamma\sqrt{\log(n)/n} \implies \|\psi_1^{\mathfrak{B}_{j\ell}}\| \le \frac{3}{2}\Gamma\sqrt{\log(n)/n}$.

PROOF. Before proving the items, we first remark that we never have $\hat{\phi}_1^2 = 1$ nor $\hat{m}_1 = 0$ on the event Ω_n thanks to Lemma 19.

We establish Item 1. Notice that

$$\begin{aligned}
\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{T}_n \iff & \left\| \frac{1 \mp \hat{\phi}_1}{1 \pm \hat{\phi}_1} \hat{\psi}_1^{\mathfrak{B}_{j\ell}} \mp \frac{\hat{g}(1 \mp \hat{\phi}_1)}{2\hat{m}_1} \hat{G}^{\mathfrak{B}_{j\ell}} \right\| > \Gamma \hat{T}_n \\
\iff & \left\| \frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \hat{\psi}_1^{\mathfrak{B}_{j\ell}} \mp \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}(1 \mp \tilde{s}\phi_1)}{2\hat{m}_1} \hat{G}^{\mathfrak{B}_{j\ell}} \right\| > \frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \frac{1 \pm \hat{\phi}_1}{1 \mp \hat{\phi}_1} \Gamma \hat{T}_n \\
\implies & \left\| \frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \hat{\psi}_1^{\mathfrak{B}_{j\ell}} \mp \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}(1 \mp \tilde{s}\phi_1)}{2\hat{m}_1} \hat{G}^{\mathfrak{B}_{j\ell}} \right\| > \frac{1}{16} \Gamma T_n
\end{aligned}$$

ie.

$$\begin{aligned}
&\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{T}_n \implies \\
&\qquad \|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\| > \frac{1}{16} \Gamma T_n - \frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| - \left\| \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}}{\hat{m}_1} \hat{G}^{\mathfrak{B}_{j\ell}} - \frac{g}{m_1} G^{\mathfrak{B}_{j\ell}} \right\|
\end{aligned}$$

where we have used Lemmas 18 and 20. But on the event $E_{j\ell} \cap \Xi_n \cap \Omega_n$

$$\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| \leq \frac{(1 \mp s\tilde{\phi}_1)^2}{1 - \phi_1^2} \cdot c_0 \Gamma \sqrt{\log(n)/n} \leq c_0 \Gamma T_n$$

and

$$\begin{aligned}
&\left\| \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}}{\hat{m}_1} \hat{G}^{\mathfrak{B}_{j\ell}} - \frac{g}{m_1} G^{\mathfrak{B}_{j\ell}} \right\| \\
&\leq \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}}{|\hat{m}_1|} \|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| + \left| \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}}{|\hat{m}_1|} - \frac{g}{m_1} \right| \|G^{\mathfrak{B}_{j\ell}}\| \\
&\leq \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}}{|\hat{m}_1|} \|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| \\
&\qquad + \left(\frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{|\hat{g} - g|}{|\hat{m}_1|} + \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{g|\hat{m}_1 - m_1|}{|\hat{m}_1 m_1|} + \frac{g}{|m_1|} \left| \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} - 1 \right| \right) \|G^{\mathfrak{B}_{j\ell}}\| \\
&\leq \frac{8g}{|m_1|} \|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| + \left(\frac{4|\hat{g} - g|}{|m_1|} + \frac{4g|\hat{m}_1 - m_1|}{m_1^2} + \frac{g|\hat{\phi}_1 - \tilde{s}\phi_1|}{(1 - \phi_1^2)|m_1|} \right) \|G^{\mathfrak{B}_{j\ell}}\|
\end{aligned}$$

where the last line holds true on Ω_n by Lemmas 11 and 18. Therefore by Lemmas 12 and 16, there is a universal constant $C > 0$ such that

$$\begin{aligned}
&\left\| \frac{1 \pm \hat{\phi}_1}{1 \pm \tilde{s}\phi_1} \frac{\hat{g}}{\hat{m}_1} \hat{G}^{\mathfrak{B}_{j\ell}} - \frac{g}{m_1} G^{\mathfrak{B}_{j\ell}} \right\| \\
&\quad \leq \frac{8g}{|m_1|} \|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| + \frac{C \max(1, g)}{|m_1 m_2|} \max_{j=1,2,3} |\hat{m}_j - m_j| \|G^{\mathfrak{B}_{j\ell}}\| \leq (8c_1 + Cc_2) \Gamma T_n
\end{aligned}$$

on the event $E_{j\ell} \cap \Xi_n \cap \Omega_n$ by definitions of these events. Therefore by choosing c_0, c_1, c_2 small enough, the Item 1 claim follows. The proof of the Item 2 is nearly identical. Items 3 and 4 are immediate from the definition of Ξ_n provided $c_0 \leq 1/2$. $\square$

In the next we make use of the symbol $\lesssim$ to denote inequalities that are valid up to a universal multiplicative constant. Furthermore, since $\hat{m}_1 \neq 0$ and $\hat{\phi}_1^2 \neq 1$ on the event Ω_n thanks to Lemma 19, and since all the terms we wish to control are conditional on Ω_n, we will assume throughout the rest of the proof that $\hat{m}_1 \neq 0$ and $\hat{\phi}_1^2 \neq 1$ without justification.

A.5.4. *Control of R_1.* This has already been done in Section A.4.1. We recall the result:

$$\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} R_1(\theta) \leq \frac{BL^2}{\delta^2 \epsilon^2 \zeta^2} \frac{\log(n)}{n\gamma^*} + \frac{BL^3}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{n\gamma^*} + \frac{B \max(\tau, L)^6}{\delta^2 \epsilon^4 \zeta^4} \frac{1}{(n\gamma^*)^2}.$$

A.5.5. *Control of R_2.*

$$\begin{aligned}
&\|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} \mathbf{1}_{\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| > \Gamma\sqrt{\log(n)/n}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} \mathbf{1}_{\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{T}_n} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| \\
&\quad = \|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}} - \hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} \mathbf{1}_{\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| \leq \Gamma\sqrt{\log(n)/n}} - \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} \mathbf{1}_{\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| \leq \Gamma \hat{T}_n}\| \\
&\quad \leq \|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| + \frac{2\Gamma\sqrt{\log(n)/n}}{1 \pm \hat{\phi}_1} + \Gamma \hat{T}_n \\
&\quad \leq \|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| + 8\Gamma T_n
\end{aligned}$$

on the event Ω_n by Lemmas 18 and 20. Furthermore, letting $\hat{f}_{\pm}^{\mathfrak{B}_{j\ell}}$ and $f_{\pm}^{\mathfrak{B}_{j\ell}}$ as defined in Section A.4, it is easily seen that

$$\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}} = \hat{f}_{\pm}^{\mathfrak{B}_{j\ell}} - f_{\pm}^{\mathfrak{B}_{j\ell}}.$$

Hence by Lemma 15, on the event $\Xi_n \cap \Omega_n$,

$$\begin{aligned}
\|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| &\leq c_0 \Gamma \sqrt{\log(n)/n} + \frac{4g}{|m_1|} c_1 \Gamma \sqrt{\log(n)/n} + \frac{1}{2} |\hat{\omega}_{\pm} - \omega_{\pm}| \|G^{\mathfrak{B}_{j\ell}}\| \\
&\leq (c_0 + 4c_1)\Gamma T_n + \frac{1}{2} |\hat{\omega}_{\pm} - \omega_{\pm}| \|G^{\mathfrak{B}_{j\ell}}\| \\
&\leq (c_0 + 4c_1)\Gamma T_n + \frac{41.5 \max(1, g)}{|m_1 m_2|} \max_{j=1,2,3} |\hat{m}_j - m_j| \|G^{\mathfrak{B}_{j\ell}}\| \qquad (22)
\end{aligned}$$

where the last line follows by Proposition 10. Deduce from the definition of $E_{j\ell}$ that on the event $E_{j\ell}^c \cap \Xi_n \cap \Omega_n$ we must have

$$\begin{aligned}
&\|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} \mathbf{1}_{\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| > \Gamma\sqrt{\log(n)/n}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} \mathbf{1}_{\|\hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}}\| > \Gamma \hat{T}_n} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| \\
&\qquad \leq \Big(\frac{8 + c_0 + 4c_1}{c_2} + 41.5\Big) \frac{\max(1, g)}{|m_1 m_2|} \max_{j=1,2,3} |\hat{m}_j - m_j| \|G^{\mathfrak{B}_{j\ell}}\|.
\end{aligned}$$

From this we obtain the estimate

$$\begin{aligned}
R_2(\theta) &\lesssim \frac{\max(1, g)^2}{m_1^2 m_2^2} \mathbb{E}_\theta \Big(\max_{j=1,2,3} |\hat{m}_j - m_j|^2\Big) \sum_{j \geq J_n} \sum_{\ell} \|G^{\mathfrak{B}_{j\ell}}\|^2 \\
&\lesssim \frac{\max(1, g)^2}{m_2^2} \Big(\frac{C^2 L^3}{n\gamma^*} + \frac{C^2 \max(\tau, L)^6}{(n\gamma^*)^2}\Big)
\end{aligned}$$

where the last line follows from Proposition 4. Therefore we deduce from Lemma 4 that

$$\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} R_2(\theta) \lesssim \frac{1}{\delta^2 \epsilon^4 \zeta^4} \Big(\frac{L^3}{n\gamma^*} + \frac{\max(\tau, L)^6}{(n\gamma^*)^2}\Big).$$

A.5.6. *Control of* R_3. By equation (22) and the definition of $E_{j\ell}$, it is found that on the event $E_{j\ell} \cap \Xi_n \cap \Omega_n$,

$$\|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} + \hat{\beta}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| \leq (c_0 + 2c_1 + 41.5c_2)\Gamma T_n.$$

Then we deduce from Lemma 21 that

$$R_3(\theta) \lesssim \Gamma^2 T_n^2 \sum_{j=J_n}^{\tilde{j}_n} \sum_{\ell} \mathbf{1}_{\{\|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\| > \frac{1}{32}\Gamma T_n\}}.$$

Noting $\beta_{\pm} = -\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1} f_{\mp}$ and mimicking the proof in Section A.4.5, it is found that

$$\sum_{\ell} \mathbf{1}_{\{\|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\| > \frac{1}{32}\Gamma T_n\}} \leq \min\left(\frac{2^j}{N}, \left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\right)^2 \frac{R^2 2^{-2js_{\mp}}}{\Gamma^2 T_n^2}\right) \tag{23}$$

Letting $A = \sup\{0 \leq j \leq \tilde{j}_n : 2^{-j(s_{\mp}+1/2)} > \frac{\Gamma T_n}{R\sqrt{N}} \frac{1\pm\tilde{s\phi_1}}{1\mp\tilde{s}\phi_1}\}$ it is found that

$$\begin{aligned}
R_3(\theta) &\lesssim \Gamma^2 T_n^2 \sum_{j=0}^{A} \frac{2^j}{N} + \Gamma^2 T_n^2 \sum_{j>A} \left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\right)^2 \frac{R^2 2^{-2js_{\mp}}}{\Gamma^2 T_n^2} \\
&\lesssim \frac{\Gamma^2 T_n^2}{N} 2^A + \left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\right)^2 R^2 \frac{2^{-2As_{\mp}}}{2^{2s_{\mp}} - 1} \\
&\lesssim \frac{\Gamma^2 T_n^2}{N} \left(\left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\right)^2 \frac{R^2 N}{\Gamma^2 T_n^2}\right)^{1/(2s_{\mp}+1)} \\
&\quad + \left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\right)^2 R^2 \frac{1}{2^{2s_{\mp}} - 1} \left(\left(\frac{1\pm\tilde{s}\phi_1}{1\mp\tilde{s}\phi_1}\right)^2 \frac{\Gamma^2 T_n^2}{R^2 N}\right)^{2s_{\mp}/(2s_{\mp}+1)} \\
&\lesssim \frac{R^2}{\min(1, s_{\mp})} \left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\right)^{2/(2s_{\mp}+1)} \left(\frac{\Gamma^2 T_n^2}{R^2 N}\right)^{2s_{\mp}/(2s_{\mp}+1)}.
\end{aligned}$$

It follows using the definition of T_n and $\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$ together with Lemma 4 (recall that $\zeta \leq 1$ by assumption) that

$$\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} R_3(\theta) \lesssim \frac{R^2}{\min(1, s_{\mp})} \frac{1}{\delta^2} \left(\frac{\Gamma^2}{R^2 n \epsilon^2 \zeta^2}\right)^{2s_{\mp}/(2s_{\mp}+1)}.$$

A.5.7. *Control of* R_4. When $\|\psi_1^{\mathfrak{B}_{j\ell}}\| > \frac{g(1\pm\tilde{s}\phi_1)}{|m_1|} \|G^{\mathfrak{B}_{j\ell}}\|$

$$\begin{aligned}
\|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\| &= \left\|\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1} \psi_1^{\mathfrak{B}_{j\ell}} \mp \frac{g(1\mp\tilde{s}\phi_1)}{2m_1} G^{\mathfrak{B}_{j\ell}}\right\| \\
&\geq \frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1} \|\psi_1^{\mathfrak{B}_{j\ell}}\| - \frac{g(1\mp\tilde{s}\phi_1)}{2|m_1|} \|G^{\mathfrak{B}_{j\ell}}\| \\
&\geq \frac{1}{2} \frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1} \|\psi_1^{\mathfrak{B}_{j\ell}}\|.
\end{aligned}$$

Consequently,

$$\|\hat{\alpha}_{\pm}^{\mathfrak{B}_{j\ell}} - \alpha_{\pm}^{\mathfrak{B}_{j\ell}} - \beta_{\pm}^{\mathfrak{B}_{j\ell}}\| = \left\|\frac{2}{1\pm\hat{\phi}_1} \hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \frac{2}{1\pm\tilde{s}\phi_1} \psi_1^{\mathfrak{B}_{j\ell}} + \left(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1} \psi_1^{\mathfrak{B}_{j\ell}} \mp \frac{g(1\mp\tilde{s}\phi_1)}{2m_1} G^{\mathfrak{B}_{j\ell}}\right)\right\|$$

$$= \Big\| \frac{2(\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}})}{1 \pm \hat{\phi}_1} + \Big(\frac{1 \mp \hat{\phi}_1}{1 \pm \hat{\phi}_1} \psi_1^{\mathfrak{B}_{j\ell}} \mp \frac{g(1 \mp \tilde{s}\phi_1)}{2m_1} G^{\mathfrak{B}_{j\ell}} \Big) \Big\|$$

$$\leq \frac{2\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\|}{1 \pm \hat{\phi}_1} + \frac{1 \mp \hat{\phi}_1}{1 \pm \hat{\phi}_1} \|\psi_1^{\mathfrak{B}_{j\ell}}\| + \frac{g(1 \mp \tilde{s}\phi_1)}{2|m_1|} \|G^{\mathfrak{B}_{j\ell}}\|$$

Then on the event $\Xi_n \cap \Omega_n$, by Lemma 18

$$\begin{aligned}
\|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}} - \alpha_\pm^{\mathfrak{B}_{j\ell}} - \beta_\pm^{\mathfrak{B}_{j\ell}}\| &\leq \frac{4c_0\Gamma\sqrt{\log(n)/n}}{1 \pm \tilde{s}\phi_1} + \frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \Big(4\|\psi_1^{\mathfrak{B}_{j\ell}}\| + \frac{g(1 \pm \tilde{s}\phi_1)}{2|m_1|} \|G^{\mathfrak{B}_{j\ell}}\| \Big) \\
&\leq \frac{4c_0\Gamma\sqrt{\log(n)/n}}{1 \pm \tilde{s}\phi_1} + 5 \frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} \|\psi_1^{\mathfrak{B}_{j\ell}}\| \\
&\leq \frac{4c_0\Gamma\sqrt{\log(n)/n}}{1 \pm \tilde{s}\phi_1} + 10\|\beta_\pm^{\mathfrak{B}_{j\ell}}\|.
\end{aligned}$$

Deduce from Lemma 21 that

$$R_4(\theta) \lesssim \frac{\Gamma^2 \log(n)/n}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \mathbf{1}_{\{\|\psi_1^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma\sqrt{\log(n)/n}\}} + \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \leq 32\Gamma T_n\}}.$$

Observe that $2\psi_1 = (1 + \tilde{s}\phi_1) f_+ + (1 - \tilde{s}\phi_1) f_-$. Therefore, for all $j \geq J_n$

$$\begin{aligned}
\sum_k |\psi_1^{\Psi_{jk}}|^2 &\leq \frac{(1 + \tilde{s}\phi_1)^2}{2} \sum_k |f_+^{\Psi_{jk}}|^2 + \frac{(1 - \tilde{s}\phi_1)^2}{2} \sum_k |f_-^{\Psi_{jk}}|^2 \\
&\leq R^2 \frac{(1 + \tilde{s}\phi_1)^2 2^{-2js_+} + (1 - \tilde{s}\phi_1)^2 2^{-2js_-}}{2}, \qquad (24)
\end{aligned}$$

whenever $\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$ (recall equation (10)). Deduce that (see also Section A.4.5)

$$\begin{aligned}
\sum_\ell \mathbf{1}_{\{\|\psi_1^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma\sqrt{\log(n)/n}\}} &\leq \min\left(\frac{2^j}{N}, \frac{2nR^2\big((1 + \tilde{s}\phi_1)^2 2^{-2js_+} + (1 - \tilde{s}\phi_1)^2 2^{-2js_-}\big)}{\Gamma^2 \log(n)} \right) \\
&\leq \frac{1}{2} \min\Big(\frac{2^j}{N}, \frac{4nR^2(1 + \tilde{s}\phi_1)^2 2^{-2js_+}}{\Gamma^2 \log(n)} \Big) \\
&\quad + \frac{1}{2} \min\Big(\frac{2^j}{N}, \frac{4nR^2(1 - \tilde{s}\phi_1)^2 2^{-2js_-}}{\Gamma^2 \log(n)} \Big)
\end{aligned}$$

by convexity of $x \mapsto \min(2^j/N, x)$. Deduce that,

$$\begin{aligned}
&\frac{\Gamma^2 \log(n)/n}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \mathbf{1}_{\{\|\psi_1^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma\sqrt{\log(n)/n}\}} \\
&\lesssim \frac{\Gamma^2}{n(1 \pm \tilde{s}\phi_1)^2} \Big(\frac{nR^2(1 + \tilde{s}\phi_1)^2}{\Gamma^2} \Big)^{1/(2s_+ + 1)} \\
&\quad + \frac{1}{2^{2s_+} - 1} \frac{R^2(1 + \tilde{s}\phi_1)^2}{(1 \pm \tilde{s}\phi_1)^2} \Big(\frac{\Gamma^2}{nR^2(1 + \tilde{s}\phi_1)^2} \Big)^{2s_+/(2s_+ + 1)} \\
&\quad + \frac{\Gamma^2}{n(1 \pm \tilde{s}\phi_1)^2} \Big(\frac{nR^2(1 - \tilde{s}\phi_1)^2}{\Gamma^2} \Big)^{1/(2s_- + 1)} \\
&\quad + \frac{1}{2^{2s_-} - 1} \frac{R^2(1 - \tilde{s}\phi_1)^2}{(1 \pm \tilde{s}\phi_1)^2} \Big(\frac{\Gamma^2}{nR^2(1 - \tilde{s}\phi_1)^2} \Big)^{2s_-/(2s_- + 1)}.
\end{aligned}$$

That is,

$$
\begin{aligned}
\frac{\Gamma^2 \log(n)/n}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \sum_{\ell} \mathbf{1}_{\{\|\psi_1^{\mathfrak{B}_{j\ell}}\| > \frac{1}{2}\Gamma\sqrt{\log(n)/n}\}}
&\lesssim \frac{R^2}{\min(1, s_+)} \Big(\frac{1 + \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 \Big(\frac{\Gamma^2}{nR^2(1 + \tilde{s}\phi_1)^2}\Big)^{2s_+/(2s_+ + 1)} \\
&\quad + \frac{R^2}{\min(1, s_-)} \Big(\frac{1 - \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 \Big(\frac{\Gamma^2}{nR^2(1 - \tilde{s}\phi_1)^2}\Big)^{2s_-/(2s_- + 1)}.
\end{aligned}
$$

Regarding the remaining term, recall that $\beta_\pm = -\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1} f_\mp$ and observe that

$$
\begin{aligned}
\sum_{j=J_n}^{\tilde{J}_n} \sum_{\ell} \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\{\|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \leq 32\Gamma T_n\}}
&\lesssim \sum_{j=J_n}^{\tilde{J}_n} \min\left(\sum_{\ell} \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2, \frac{2^j \Gamma^2 T_n^2}{N} \right) \\
&\lesssim \sum_{j=J_n}^{\tilde{J}_n} \min\left(\sum_{\ell} \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2, \frac{2^j \Gamma^2 T_n^2}{N} \right) \\
&\lesssim \sum_{j=J_n}^{\tilde{J}_n} \min\left(R^2 \Big(\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 2^{-2js_\mp}, \frac{2^j \Gamma^2 T_n^2}{N} \right) \\
&\lesssim \frac{R^2}{\min(1, s_\mp)} \Big(\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^{2/(2s_\mp + 1)} \Big(\frac{\Gamma^2 T_n^2}{R^2 N}\Big)^{2s_\mp/(2s_\mp + 1)} \qquad (25)
\end{aligned}
$$

where the last line follows from the estimate in (23) and subsequent iterates. In the end,

$$
\begin{aligned}
R_4(\theta) &\lesssim \frac{R^2}{\min(1, s_+)} \Big(\frac{1 + \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 \Big(\frac{\Gamma^2}{nR^2(1 + \tilde{s}\phi_1)^2}\Big)^{2s_+/(2s_+ + 1)} \\
&\quad + \frac{R^2}{\min(1, s_-)} \Big(\frac{1 - \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 \Big(\frac{\Gamma^2}{nR^2(1 - \tilde{s}\phi_1)^2}\Big)^{2s_-/(2s_- + 1)} \\
&\quad + \frac{R^2}{\min(1, s_\mp)} \Big(\frac{1 \mp \tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^{2/(2s_\mp + 1)} \Big(\frac{\Gamma^2 T_n^2}{R^2 N}\Big)^{2s_\mp/(2s_\mp + 1)}.
\end{aligned}
$$

Taking the suprema of each terms, with the help of Lemma 4 it is found that

$$
\begin{aligned}
\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} R_4(\theta) &\lesssim \frac{R^2}{\min(1, s_\pm)} \Big(\frac{\Gamma^2}{nR^2\delta^2}\Big)^{2s_\pm/(2s_\pm + 1)} \\
&\quad + \frac{R^2}{\min(1, s_\mp)} \frac{1}{\delta^2} \Big(\frac{\Gamma^2}{nR^2}\Big)^{2s_\mp/(2s_\mp + 1)} \\
&\quad + \frac{R^2}{\min(1, s_\mp)} \frac{1}{\delta^2} \Big(\frac{\Gamma^2}{R^2 n \epsilon^2 \zeta^2}\Big)^{2s_\mp/(2s_\mp + 1)}.
\end{aligned}
$$

Namely,

$$
\begin{aligned}
\sup_{\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)} R_4(\theta) &\lesssim \frac{R^2}{\min(1, s_\pm)} \Big(\frac{\Gamma^2}{nR^2\delta^2}\Big)^{2s_\pm/(2s_\pm + 1)} \\
&\quad + \frac{R^2}{\min(1, s_\mp)} \frac{1}{\delta^2} \Big(\frac{\Gamma^2}{R^2 n \epsilon^2 \zeta^2}\Big)^{2s_\mp/(2s_\mp + 1)}.
\end{aligned}
$$

A.5.8. *Control of* R_5. When $\|\psi_1^{\mathfrak{B}_{j\ell}}\| \leq \frac{g(1\pm\tilde{s}\phi_1)}{|m_1|}\|G^{\mathfrak{B}_{j\ell}}\|$,

$$\begin{aligned}
\|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}} - \alpha_\pm^{\mathfrak{B}_{j\ell}} - \beta_\pm^{\mathfrak{B}_{j\ell}}\| &\leq \|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}} - \alpha_\pm^{\mathfrak{B}_{j\ell}}\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&= \Big\|\frac{2}{1\pm\hat{\phi}_1}\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \frac{2}{1\pm\tilde{s}\phi_1}\psi_1^{\mathfrak{B}_{j\ell}}\Big\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\leq \frac{2}{1\pm\hat{\phi}_1}\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| + 2\|\psi_1^{\mathfrak{B}_{j\ell}}\|\Big|\frac{1}{1\pm\hat{\phi}_1} - \frac{1}{1\pm\tilde{s}\phi_1}\Big| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\leq \frac{2}{1\pm\hat{\phi}_1}\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}} - \psi_1^{\mathfrak{B}_{j\ell}}\| + 2\|\psi_1^{\mathfrak{B}_{j\ell}}\|\frac{|\hat{\phi}_1 - \tilde{s}\phi_1|}{(1\pm\hat{\phi}_1)(1\pm\tilde{s}\phi_1)} + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|
\end{aligned}$$

So by Lemmas 12 and 18, it holds on the event $E_{j\ell}\cap\Xi_n\cap\Omega_n$

$$\begin{aligned}
&\|\hat{\alpha}_\pm^{\mathfrak{B}_{j\ell}} - \alpha_\pm^{\mathfrak{B}_{j\ell}} - \beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\quad\leq \frac{4c_0\Gamma\sqrt{\log(n)/n}}{1\pm\tilde{s}\phi_1 l} + \frac{800\max(1,g)}{\phi_2^2\phi_3^2\tilde{\mathcal{I}}^2 g}\frac{\max_{j=1,2,3}|\hat{m}_j - m_j|}{(1\pm\tilde{s}\phi_1)^2}\|\psi_1^{\mathfrak{B}_{j\ell}}\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\quad\leq \frac{4c_0\Gamma\sqrt{\log(n)/n}}{1\pm\tilde{s}\phi_1} + \frac{800\max(1,g)}{\phi_2^2\phi_3^2\tilde{\mathcal{I}}^2}\frac{\max_{j=1,2,3}|\hat{m}_j - m_j|}{|m_1|(1\pm\tilde{s}\phi_1)}\|G^{\mathfrak{B}_{j\ell}}\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\quad\leq \frac{4c_0\Gamma\sqrt{\log(n)/n}}{1\pm\tilde{s}\phi_1} + \frac{800\max(1,g)}{|m_1m_2|}\max_{j=1,2,3}|\hat{m}_j - m_j|\|G^{\mathfrak{B}_{j\ell}}\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|.
\end{aligned}$$

From here, it is seen that an upper bound on the supremum of R_5 is obtained by adding the bounds obtained on R_2 together with the bound on R_4, eventually up to a universal multiplicative constant.

A.5.9. *Control of* R_6.

$$\begin{aligned}
&\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}} - \alpha_\pm^{\mathfrak{B}_{j\ell}} - \beta_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\quad\leq \|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}} - \beta_\pm^{\mathfrak{B}_{j\ell}}\| + \|\alpha_\pm^{\mathfrak{B}_{j\ell}}\| \\
&\quad= \Big\|\frac{1\mp\hat{\phi}_1}{1\pm\hat{\phi}_1}\hat{\psi}_1^{\mathfrak{B}_{j\ell}} \mp \frac{\hat{g}(1\mp\hat{\phi}_1)}{2\hat{m}_1}\hat{G}^{\mathfrak{B}_{j\ell}} - \Big(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\psi_1^{\mathfrak{B}_{j\ell}} \mp \frac{g(1\mp\tilde{s}\phi_1)}{2m_1}G^{\mathfrak{B}_{j\ell}}\Big)\Big\| \\
&\qquad+ \frac{2}{1\pm\tilde{s}\phi_1}\|\psi_1^{\mathfrak{B}_{j\ell}}\| \\
&\quad\leq \frac{3}{1\pm\tilde{s}\phi_1}\|\psi_1^{\mathfrak{B}_{j\ell}}\| + \frac{1}{1\pm\hat{\phi}_1}\|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\| + \Big\|\frac{\hat{g}(1\mp\hat{\phi}_1)}{2\hat{m}_1}\hat{G}^{\mathfrak{B}_{j\ell}} - \frac{g(1\mp\tilde{s}\phi_1)}{2m_1}G^{\mathfrak{B}_{j\ell}}\Big\|
\end{aligned}$$

but by Proposition 10 on the event Ω_n we have

$$\begin{aligned}
\Big\|\frac{\hat{g}(1\mp\hat{\phi}_1)}{2\hat{m}_1}\hat{G}^{\mathfrak{B}_{j\ell}} - \frac{g(1\mp\tilde{s}\phi_1)}{2m_1}G^{\mathfrak{B}_{j\ell}}\Big\| &= |\hat{\omega}_\mp|\|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| + |\hat{\omega}_\mp - \omega_\mp|\|G^{\mathfrak{B}_{j\ell}}\| \\
&\lesssim \frac{g}{|m_1|}\|\hat{G}^{\mathfrak{B}_{j\ell}} - G^{\mathfrak{B}_{j\ell}}\| \\
&\qquad+ \frac{\max(1,g)}{|m_1m_2|}\max_{j=1,2,3}|\hat{m}_j - m_j|\|G^{\mathfrak{B}_{j\ell}}\|.
\end{aligned}$$

Therefore on the event $E_{j\ell} \cap \Xi_n \cap \Omega_n$

$$\|\hat{\beta}_\pm^{\mathfrak{B}_{j\ell}} - \alpha_\pm^{\mathfrak{B}_{j\ell}} - \beta_\pm^{\mathfrak{B}_{j\ell}}\| \lesssim \frac{\|\psi_1^{\mathfrak{B}_{j\ell}}\| + \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|}{1 \pm \tilde{s}\phi_1} + \frac{g}{|m_1|} c_1 \Gamma \sqrt{\log(n)/n} + c_2 \Gamma T_n$$
$$\leq \frac{\|\psi_1^{\mathfrak{B}_{j\ell}}\| + \|\hat{\psi}_1^{\mathfrak{B}_{j\ell}}\|}{1 \pm \tilde{s}\phi_1} + (c_1 + c_2)\Gamma T_n.$$

Deduce by Lemma 21 that

$$R_6(\theta) \lesssim \Gamma^2 T_n^2 \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \mathbf{1}_{\|\beta_\pm^{\mathfrak{B}_{j\ell}}\| > \frac{1}{32}\Gamma T_n}.$$

Therefore, $R_6(\theta)$ admits the same upper bound as $R_3(\theta)$, eventually up to a universal multiplicative factor.

A.5.10. *Control of R_7.*

$$\|\alpha_\pm^{\mathfrak{B}_{j\ell}} + \beta_\pm^{\mathfrak{B}_{j\ell}}\| \leq \|\alpha_\pm^{\mathfrak{B}_{j\ell}}\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\| = \frac{2}{1 \pm \tilde{s}\phi_1} \|\psi_1^{\mathfrak{B}_{j\ell}}\| + \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|$$

Therefore, we obtain from Lemma 21 that

$$R_7(\theta) \leq \frac{2}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\psi_1^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\|\psi_1^{\mathfrak{B}_{j\ell}}\| \leq \frac{3}{2}\Gamma\sqrt{\log(n)/n}}$$
$$+ 2 \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\beta_\pm^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\|\beta_\pm^{\mathfrak{B}_{j\ell}}\| \leq 32\Gamma T_n}$$

From equation (24),

$$\frac{2}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\psi_1^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\|\psi_1^{\mathfrak{B}_{j\ell}}\| \leq \frac{3}{2}\Gamma\sqrt{\log(n)/n}}$$
$$\leq \frac{2}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \min\left(\frac{9\Gamma^2 \log(n)}{4n} \frac{2^j}{N}, \sum_\ell \|\psi_1^{\mathfrak{B}_{j\ell}}\|^2 \right)$$
$$\lesssim \frac{1}{(1 \pm \tilde{s}\phi_1)^2} \frac{\Gamma^2 \log(n)}{n} \sum_{j=J_n}^{\tilde{J}_n} \min\left(\frac{2^j}{N}, nR^2 \frac{(1+\tilde{s}\phi_1)^2 2^{-2js_+} + (1-\tilde{s}\phi_1)^2 2^{-2js_-}}{\Gamma^2 \log(n)} \right)$$

Then deduce from the series of estimates after (24) that

$$\frac{2}{(1 \pm \tilde{s}\phi_1)^2} \sum_{j=J_n}^{\tilde{J}_n} \sum_\ell \|\psi_1^{\mathfrak{B}_{j\ell}}\|^2 \mathbf{1}_{\|\psi_1^{\mathfrak{B}_{j\ell}}\| \leq \frac{3}{2}\Gamma\sqrt{\log(n)/n}}$$
$$\lesssim \frac{R^2}{\min(1, s_+)} \Big(\frac{1+\tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 \Big(\frac{\Gamma^2}{nR^2(1+\tilde{s}\phi_1)^2}\Big)^{2s_+/(2s_+ + 1)}$$
$$+ \frac{R^2}{\min(1, s_-)} \Big(\frac{1-\tilde{s}\phi_1}{1 \pm \tilde{s}\phi_1}\Big)^2 \Big(\frac{\Gamma^2}{nR^2(1-\tilde{s}\phi_1)^2}\Big)^{2s_-/(2s_- + 1)}.$$

Next, it has been already established in (25) that

$$\sum_{j=J_n}^{\tilde{j}_n}\sum_{\ell}\|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\|^2\mathbf{1}_{\|\beta_{\pm}^{\mathfrak{B}_{j\ell}}\|\le 32\Gamma T_n}\lesssim\frac{R^2}{\min(1,s_{\mp})}\Big(\frac{1\mp\tilde{s}\phi_1}{1\pm\tilde{s}\phi_1}\Big)^{2/(2s_{\mp}+1)}\Big(\frac{\Gamma^2T_n^2}{R^2N}\Big)^{2s_{\mp}/(2s_{\mp}+1)}.$$

Consequently, when passing to the supremum, R_7 will obey the same upper bound as R_4, eventually up to a universal multiplicative constant.

A.5.11. *Control of R_8.* This has already been done in Section A.4.6. We recall the result:

$$\sup_{\theta\in\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)\cap\Sigma_{\gamma^*}(L)}R_8(\theta)\le\frac{BR^2}{\min(1,s_{\pm})}\Big(\frac{\tau^2\log(n)}{n}\Big)^{2s_{\pm}}.$$

A.6. Proof of Theorem 6. Recall $\tilde{V}$ is the leading eigenvector of the empirical Gram matrix $\tilde{\mathcal{G}}$ and V_θ the leading eigenvector of the Gram matrix $\mathcal{G}$ normalized such that $\|\tilde{V}\|=\|V_\theta\|=1$. We use a Davis-Kahan argument to bound the norm $\|\tilde{V}-\mathrm{sgn}(\langle\tilde{V},V_\theta\rangle)V_\theta\|$. In particular using the version of Davis-Kahan's theorem given in the Corollary 1 of [20], we know that

$$\|\tilde{V}-\mathrm{sgn}(\langle\tilde{V},V_\theta\rangle)V_\theta\|\le\frac{2\sqrt{2}\|\tilde{\mathcal{G}}-\mathcal{G}\|_{\mathrm{op}}}{|\lambda|}$$

where λ is the unique non-zero eigenvalue of $\mathcal{G}$, and $\|\cdot\|_{\mathrm{op}}$ stands for the operator norm. It is rapidly seen that

$$\lambda=r(\phi)\sum_{\lambda\in\Lambda(M)}\langle\psi_2,e_\lambda\rangle^2=r(\phi)\Bigg(\sum_{k=0}^{2^J-1}\langle\psi_2,\Phi_{Jk}\rangle^2+\sum_{j=J}^{M}\sum_{k=0}^{2^j-1}\langle\psi_2,\Psi_{jk}\rangle^2\Bigg).$$

We now bound $\|\tilde{\mathcal{G}}-\mathcal{G}\|_{\mathrm{op}}$. By definition of the operator norm and then by a duality argument [here U denotes the unit ball of $\mathbb{R}^{\Lambda(M)}$]

$$\begin{aligned}\|\tilde{\mathcal{G}}-\mathcal{G}\|_{\mathrm{op}}&=\sup_{u\in U}\|\tilde{\mathcal{G}}u-\mathcal{G}u\|\\&=\sup_{u\in U}\sup_{v\in U}v^T(\tilde{\mathcal{G}}-\mathcal{G})u\\&=\sup_{u\in U}\sup_{v\in U}\Big[\Big(\frac{u+v}{2}\Big)^T(\tilde{\mathcal{G}}-\mathcal{G})\frac{u+v}{2}-\Big(\frac{u-v}{2}\Big)^T(\tilde{\mathcal{G}}-\mathcal{G})\frac{u-v}{2}\Big]\\&\le\sup_{u\in U}\sup_{v\in U}\Big[u^T(\tilde{\mathcal{G}}-\mathcal{G})u-v^T(\tilde{\mathcal{G}}-\mathcal{G})v\Big]\\&\le 2\sup_{u\in U}u^T(\tilde{\mathcal{G}}-\mathcal{G})u.\end{aligned}$$

Then, let $\mathcal{N}$ be a $(1/8)$-net over U in the euclidean norm, and let $\pi:U\to\mathcal{N}$ denote the map that projects elements of U onto their closest element in $\mathcal{N}$. Then,

$$\begin{aligned}\sup_{u\in U}u^T(\tilde{\mathcal{G}}-\mathcal{G})u&=\sup_{u\in U}\Big[\pi(u)^T(\tilde{\mathcal{G}}-\mathcal{G})\pi(u)+2\pi(u)^T(\tilde{\mathcal{G}}-\mathcal{G})(u-\pi(u))\\&\qquad+(u-\pi(u))^T(\tilde{\mathcal{G}}-\mathcal{G})(u-\pi(u))\Big]\\&\le\max_{u\in\mathcal{N}}u^T(\tilde{\mathcal{G}}-\mathcal{G})u+\frac{3}{8}\|\tilde{\mathcal{G}}-\mathcal{G}\|_{\mathrm{op}}\end{aligned}$$

and thus

$$\|\tilde{\mathcal{G}} - \mathcal{G}\|_{\text{op}} \leq 8 \max_{u \in \mathcal{N}} u^T(\tilde{\mathcal{G}} - \mathcal{G})u.$$

Next, we decompose $\tilde{\mathcal{G}} - \mathcal{G} = \Delta^{(1)} + \Delta^{(2)} + \Delta^{(3)} + \Delta^{(4)}$ with

$$\begin{aligned}
\Delta^{(1)}_{\lambda\lambda'} &:= \frac{1}{2}\Big(\tilde{\mathbb{P}}^{(1)}_n(e_\lambda \otimes e_{\lambda'} + e_{\lambda'} \otimes e_\lambda) - \mathbb{E}_\theta(e_\lambda \otimes e_{\lambda'} + e_{\lambda'} \otimes e_\lambda)\Big)\\
\Delta^{(2)}_{\lambda\lambda'} &:= -\mathbb{E}_\theta(e_{\lambda'})\Big(\tilde{\mathbb{P}}^{(1)}_n(e_\lambda) - \mathbb{E}_\theta(e_\lambda)\Big)\\
\Delta^{(3)}_{\lambda\lambda'} &:= -\mathbb{E}_\theta(e_\lambda)\Big(\tilde{\mathbb{P}}^{(1)}_n(e_{\lambda'}) - \mathbb{E}_\theta(e_{\lambda'})\Big)\\
\Delta^{(4)}_{\lambda\lambda'} &:= -\Big(\tilde{\mathbb{P}}^{(1)}_n(e_\lambda) - \mathbb{E}_\theta(e_\lambda)\Big)\Big(\tilde{\mathbb{P}}^{(1)}_n(e_{\lambda'}) - \mathbb{E}_\theta(e_{\lambda'})\Big)
\end{aligned}$$

Using Lemma 7 applied to the function $h(y_1, y_2) = \frac{1}{2}\sum_{\lambda,\lambda' \in \Lambda(M)} u_\lambda u_{\lambda'}\big(e_\lambda(y_1)e_{\lambda'}(y_2) + e_{\lambda'}(y_1)e_\lambda(y_2)\big)$ we find that

$$\begin{aligned}
\mathbb{P}_\theta\Big(\max_{u \in \mathcal{N}} |u^T\Delta^{(1)}u| \geq x\Big) &\leq |\mathcal{N}| \max_{u \in |\mathcal{N}|} \mathbb{P}_\theta\Big(|u^T\Delta^{(1)}u| \geq x\Big)\\
&\leq 24^{2^M} \exp\left(-\frac{Cn\gamma^* x^2}{L^2 + 2^M x}\right)
\end{aligned}$$

because $\mathcal{N}$ can always be chosen to have cardinality no more than 24^{2^M} (e.g. [12, Theorem 4.3.34]), because $\mathbb{E}_\theta(h^2) \leq L^2\|h\|^2_{L^2} = L^2$ for all $\theta \in \Sigma_{\gamma^*}(L)$ by Lemma 5, and because

$$\begin{aligned}
\|h\|_\infty &\leq \sup_{y_1,y_2}\Big|\sum_{\lambda \in \Lambda(M)} u_\lambda e_\lambda(y_1) \sum_{\lambda' \in \Lambda(M)} u_{\lambda'} e_{\lambda'}(y_2)\Big|\\
&\leq \Big(\sup_y \sum_{\lambda \in \Lambda(M)} \big|e_\lambda(y)\big|\Big)^2\\
&\leq c2^M
\end{aligned}$$

for a constant $c > 0$ depending only on the wavelet basis by a standard localization properties of wavelets [12, Theorem 4.2.10 or Definition 4.2.14]. Next, note that

$$u^T\Delta^{(2)}u = u^T\Delta^{(3)}u = -\mathbb{E}_\theta\left(\sum_{\lambda \in \Lambda(M)} u_\lambda e_\lambda\right)\left(\sum_{\lambda \in \Lambda(M)} u_\lambda\Big(\tilde{\mathbb{P}}^{(1)}_n(e_\lambda) - \mathbb{E}_\theta(e_\lambda)\Big)\right)$$

and,

$$u^T\Delta^{(4)}u = -\left(\sum_{\lambda \in \Lambda(M)} u_\lambda\Big(\tilde{\mathbb{P}}^{(1)}_n(e_\lambda) - \mathbb{E}_\theta(e_\lambda)\Big)\right)^2.$$

Again using Lemma 7, this time applied to the function $h(y) = \sum_{\lambda \in \Lambda(M)} u_\lambda e_\lambda(y)$ which satisfies $\mathbb{E}_\theta(h^2) \leq L$ for all $\theta \in \Sigma_{\gamma^*}(L)$ and $\|h\|_\infty \leq c2^{M/2}$ for a universal constant $c > 0$, we deduce that

$$\mathbb{P}_\theta\left(\max_{u \in \mathcal{N}}\Big|\sum_{\lambda \in \Lambda(M)} u_\lambda\Big(\tilde{\mathbb{P}}^{(1)}_n(e_\lambda) - \mathbb{E}_\theta(e_\lambda)\Big)\Big| \geq x\right) \leq 24^{2^M} \exp\left(-\frac{Cn\gamma^* x^2}{L + 2^{M/2}x}\right).$$

Since $|\mathbb{E}_\theta h| \le [\mathbb{E}_\theta h^2]^{1/2} \le \sqrt{L}$, using that $L, 2^{M/2} \ge 1$, we deduce that

$$\mathbb{P}_\theta\Big(\tfrac{1}{8}\|\tilde{\mathcal{G}} - \mathcal{G}\|_{\mathrm{op}} \ge (2\sqrt{L}+1)x + x^2\Big) \le 2\cdot 24^{2^M}\exp\left(-\frac{Cn\gamma^* x^2}{L^2 + 2^M x}\right)$$

for a constant $C > 0$. This entails that

$$\mathbb{P}_\theta\left(\|\tilde{V} - \mathrm{sgn}(\langle \tilde{V}, V_\theta\rangle)V_\theta\| \ge \frac{16\sqrt{2}\big((2\sqrt{L}+1)x + x^2\big)}{|r(\phi)|\sum_{\lambda\in\Lambda(M)}\langle \psi_2, e_\lambda\rangle^2}\right) \le 2\cdot 24^{2^M}\exp\left(-\frac{Cn\gamma^* x^2}{L^2 + 2^M x}\right)$$

Let us remark that the wavelets coefficients of ψ_2 are those of $(f_0 - f_1)/\phi_3$. Hence, whenever $\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$, from the definition of $\Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R)$ and of the Besov norm in equation (10) it must be that

$$\sup_{j\ge J} 2^{2js_*}\sum_{k=0}^{2^j-1}|\langle \psi_2, \Psi_{jk}\rangle|^2 \le \frac{4R^2}{\phi_3^2}, \tag{26}$$

Consequently since $\|\psi_2\|_{L^2} = 1$:

$$\begin{aligned}
1 &= \sum_{k=0}^{2^J-1}\langle \psi_2, \Phi_{Jk}\rangle^2 + \sum_{j\ge J}\sum_{k=0}^{2^j-1}\langle \psi_2, \Psi_{jk}\rangle^2 \\
&\le \sum_{k=0}^{2^J-1}\langle \psi_2, \Phi_{Jk}\rangle^2 + \sum_{j=J}^{M}\sum_{k=0}^{2^j-1}\langle \psi_2, \Psi_{jk}\rangle^2 + \frac{4R^2}{\phi_3^2}\sum_{j>M}2^{-2js_*} \\
&= \sum_{\lambda\in\Lambda(M)}\langle \psi_2, e_\lambda\rangle^2 + \frac{4R^2}{\phi_3^2}\frac{2^{-2Ms_*}}{2^{2s_*}-1}.
\end{aligned}$$

and hence $\sum_{\lambda\in\Lambda(M)}\langle \psi_2, e_\lambda\rangle^2 \ge 3/4$ under the assumptions of the theorem. Observe that $|r(\phi)| \le \phi_3^2/4 \le L/2$ by Lemmas 6 and 3. Then taking $x = \kappa|r(\phi)|/\sqrt{L}$ for a small enough constant κ, we find that for some $C > 0$

$$\mathbb{P}_\theta\left(\|\tilde{V} - \mathrm{sgn}(\langle \tilde{V}, V_\theta\rangle)V_\theta\| \ge \frac{1}{5}\right) \le 2\cdot 24^{2^M}\exp\left(-\frac{Cn\gamma^* r(\phi)^2}{L^3 + 2^M\sqrt{L}|r(\phi)|}\right).$$

Next, let define $t := \sum_{\lambda\in\Lambda(M)}\tilde{V}_\lambda e_\lambda$ and $f(x) := \max(-\tau, \min(\tau, x))$. Observe that

$$\|\psi_2\|_\infty = \frac{\|f_0 - f_1\|_\infty}{\phi_3} \le \frac{L}{\zeta}$$

since $0 \le f_0, f_1 \le L$ and $\phi_3 \ge \zeta$ when $\theta \in \Theta_{\delta,\epsilon,\zeta}^{s_0,s_1}(R) \cap \Sigma_{\gamma^*}(L)$. Then by assumption $|\psi_2(x)| \le \tau$ for all x, and thus $\psi_2(x) = f(\psi_2(x))$. Also f is 1-Lipschitz, and thus

$$\|f\circ t - \tilde{s}\psi_2\|_{L^2} = \|f\circ t - f\circ(\tilde{s}\psi_2)\|_{L^2} \le \|t - \tilde{s}\psi_2\|_{L^2} = \|\tilde{V} - \mathrm{sgn}\big(\langle \tilde{V}, V_\theta\rangle\big)V_\theta\|.$$

Since $\tilde{\psi_2} = f\circ t/\|f\circ t\|_{L^2}$, we use that for any norm $\|a/\|a\| - b/\|b\|\| \le 2\|a-b\|/(1 - \|a-b\|)$ if $\|b\| = 1$, $\|a-b\| < 1$ to deduce that

$$\|\tilde{\psi_2} - \tilde{s}\psi_2\|_{L^2} \le \frac{2\|\tilde{V} - \mathrm{sgn}\big(\langle \tilde{V}, V_\theta\rangle\big)V_\theta\|}{1 - \|\tilde{V} - \mathrm{sgn}\big(\langle \tilde{V}, V_\theta\rangle\big)V_\theta\|}.$$

The conclusion follows since $\|\tilde{\psi_2} - \tilde{s}\psi_2\|_{L^2}^2 = 2 - 2|\langle \tilde{\psi_2}, \psi_2\rangle|$, and hence $|\langle \tilde{\psi_2}, \psi_2\rangle| \ge 1 - \frac{\|\tilde{\psi_2} - \tilde{s}\psi_2\|_{L^2}^2}{2}$.

A.7. Proof of Corollary 1. In the considered regime $n\delta^2\epsilon^4\zeta^6 \geq n^{1-2a-4b-6c}$ while $L^3 + \max(\tau, \sqrt{L})^3\delta\epsilon^2\zeta^3 \leq L^3 + \max(\tau, \sqrt{L})^3$ since $\delta\epsilon^2\zeta^3 \leq 1$. Hence, the exponential term in the bound of Theorem 3 is smaller than $\exp(-Kn^{1-2a-4b-6c})$ for some $K > 0$ and is negligible.

We claim that the term $\frac{1}{\delta^2\epsilon^2\zeta^2}\frac{\log(n)}{n}$ never dominates. Indeed, for this term to dominate, it is necessary that $\epsilon^2\zeta^2 \gg \frac{1}{\log(n)}$ to dominate the term $\frac{1}{\delta^2\epsilon^4\zeta^4 n}$ and that $\delta^2\epsilon^2\zeta^2 n = O(\log(n)^{2s_i+1})$ to dominate the term $(\delta^2\epsilon^2\zeta^2 n)^{-2s_i/(2s_i+1)}$, *ie.* $\epsilon^2\zeta^2 = O(\frac{\log(n)^{2s_i+1}}{n\delta^2}) = O(\frac{\log(n)^{2s_i+1}}{n^{1-2a}})$. Since $1 - 2a > 0$, the two requirements cannot be fulfilled simultaneously for n large.

Finally, the term $\frac{1}{\delta^2\epsilon^4\zeta^6 n^2}$ is clearly dominated by the term $\frac{1}{\delta^2\epsilon^4\zeta^6 n}$ and the remaining term is clearly dominated by the term $(\delta^2\epsilon^2\zeta^2 n)^{-2s_i/(2s_i+1)}$.

A.8. Proof of Corollary 2. As for the proof of Corollary 1 the exponential term in the bound of Theorem 5 cannot dominate in the considered regime. It has been shown in Corollary 1 that the term $\frac{\log(n)}{\delta^2\epsilon^2\zeta^2 n}$ cannot simultaneously dominate the terms $\frac{1}{\delta^2\epsilon^4\zeta^4 n}$ and $\delta^{-2}(n\epsilon^2\zeta^2)^{-2s_1/(2s_1+1)}$ [observe that $\delta^{-2}(n\epsilon^2\zeta^2)^{-2s_1/(2s_1+1)} \geq (n\delta^2\epsilon^2\zeta^2)^{-2s_1/(2s_1+1)}$]. Also using the arguments in the proof of Corollary 1 it is trivial that the terms $\frac{1}{\delta^2\epsilon^4\zeta^4 n^2}$ and $(\log(n)/n)^{2s_0}$ cannot dominate.

To finish the proof, it is enough to show that the term $\delta^{-2}(n\epsilon^2\zeta^2)^{-2s_1/(2s_1+1)}$ is dominated by the term $(n\delta^2)^{-2s_0/(2s_0+1)}$. But in the considered regime $\delta^{-2}(n\epsilon^2\zeta^2)^{-2s_1/(2s_1+1)} = n^{-2s_1/(2s_1+1)+o(1)}$ and $(n\delta^2)^{-2s_0/(2s_0+1)} = n^{-2s_0/(2s_0+1)+o(1)}$. The conclusion follows since $s_1 > s_0$ by assumption.

APPENDIX B: PROOFS FOR THE LOWER BOUNDS

For proving our lower bounds, we shall follow the usual path, in which we need at some point upper bounds for distances between joint distributions $P_\theta^{(n)}$ for different values of θ. We shall use the same trick as the one used in [2], that is an upper bound on the Kullback-Leibler divergence using a pseudo-distance ρ between parameters, see the end of Section III in [2] for heuristics explaining the importance of ρ interpreted as a fundamental statistical distance in HMM learning.

The following result is Proposition 2 in [2], for which a close look at the proof shows that it still holds with emission densities on $[0, 1]$ instead of probability mass functions.

PROPOSITION 12. *Assume there exists $c > 0$ such that uniformly on $[0, 1]$ it holds $\min(f_0, f_1, \tilde{f}_0, \tilde{f}_1) \geq c$. Then*

$$K(P_\theta^{(n)}, P_{\tilde{\theta}}^{(n)}) \leq Cn\rho(\phi(\theta), \psi(\theta); \phi(\tilde{\theta}), \psi(\tilde{\theta}))^2, \tag{27}$$

where, as in [2], we have defined

$$\begin{aligned}\rho(\phi, \psi; \tilde{\phi}, \tilde{\psi}) = \max\big\{&|r(\phi) - r(\tilde{\phi})|, |\phi_2 r(\phi) - \tilde{\phi}_2 r(\tilde{\phi})|,\\ &|\phi_1\phi_2\phi_3 r(\phi) - \mathrm{sgn}\big(\langle\psi_2, \tilde{\psi}_2\rangle\big)\tilde{\phi}_1\tilde{\phi}_2\tilde{\phi}_3 r(\tilde{\phi})|,\\ &\|\psi_1 - \tilde{\psi}_1\|_{L^2}, \max(|r(\phi)|, |r(\tilde{\phi})|)\|\psi_2 - \mathrm{sgn}\big(\langle\psi_2, \tilde{\psi}_2\rangle\big)\tilde{\psi}_2\|_{L^2}\big\}.\end{aligned} \tag{28}$$

[Recall $r(\phi) = (1/4)(1 - \phi_1^2)\phi_2\phi_3^2$.]

B.1. Proof of Theorem 2. To prove Theorem 2, we shall use a standard two-points argument using Le Cam's method ([13], see also [19] for a review of lower bound ideas): if θ and $\tilde{\theta}$ in $\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)$ are such that $|p-\tilde{p}|^2\geq R_n$ and $K(P^{(n)}_\theta,P^{(n)}_{\tilde{\theta}})\leq\alpha<1$, then

$$\inf_{\hat{\theta}}\sup_{\theta\in\Theta^{s_0,s_1}_{\delta,\epsilon,\zeta}(R)\cap\Sigma_{\gamma^*}(L)}\mathbb{E}_\theta\left(|\hat{p}-p|^2\right)\geq\frac{R_n}{4}\left(1-\sqrt{\alpha}\right).$$

We follow the method in the multinomial case (see [2]) used to choose the two points in proving Theorems 1 and 3 therein, except that rather than defining ψ according to Lemma 3 therein we choose $\psi_1=1$ and $\psi_2(x)=\sqrt{3}(2x-1)$. This choice of $\tilde{\psi}=\psi$ leads to lower bounded f_0 and f_1 (so that we can apply Proposition 12) when $\|f_0-f_1\|_{L^2}=\zeta\leq1/(4\sqrt{3})$, $\|f_i\|_\infty\leq5/8$ and $\|f_i\|_{B^{s_i}_{2,\infty}}\leq5/4+1/(8\sqrt{3})$, $i=0,1$, as a consequence of the inversion formulae (Lemma 1). Under the assumption that for a suitable $\epsilon_0>0$ we have $\zeta\leq1/(4\sqrt{3})$, $\gamma^*\leq1/3$, $\epsilon\leq\epsilon_0$, $\delta\leq1/6$, the proof of the lower bounds for ϕ in Theorem 3 and the lower bound for p in Theorem 1 in [2] goes through to get the result. [Note that the bound obtained for ϕ_3 using this method is not sharp, but is also not of interest since below we lower bound estimation rates for f_0,f_1 directly without passing via ϕ_3.]

B.2. Proof of Theorem 4. For the parametric term in the lower bound, we are again able to copy the proof of [2] Theorems 1 and 3 up to the choice of ψ. Under the assumption that for a suitable $\epsilon_0>0$ we have $\zeta\leq1/(4\sqrt{3})$, $\gamma^*\leq1/3$, $\epsilon\leq\epsilon_0$, $\delta\leq1/6$, as with proving Theorem 2 we choose $\psi_1=1$, $\psi_2(x)=\sqrt{3}(2x-1)$, and the proof of the lower bound for f_0 in [2, Theorem 1] goes through.

We now prove the lower bound given in the second part of the theorem

$$R_{\text{smooth}}=(n\delta^2\epsilon^2\zeta^2)^{-s_0/(2s_0+1)}$$

We proceed via a usual reduction to multiple testing, see for instance [18]. For a suitable c,α, it suffices to construct function $f_{0,m}\in H^{s_0}(R)$, $f_{1,m}\in H^{s_1}(R)$, $0\leq m\leq M=\lceil2^{c2^j}\rceil$, for some j, such that

$$K\left(P^{(n)}_m,P^{(n)}_0\right)\leq c\alpha2^j,\quad\|f_{0,m}-f_{0,m'}\|_{L^2}\geq cR_{\text{smooth}},\tag{29}$$

where $P^{(n)}_m$ denotes the law of $(Y_1,\dots,Y_n)$ under parameter $\theta_m=(p_m,q_m,f_{0,m},f_{1,m})$ (for suitable choices of the parameters p_m,q_m). Indeed, given such functions, we note that

$$\frac{1}{M\log M}\sum_{m=1}^M K\left(P^{(n)}_m,P^{(n)}_0\right)\leq\alpha,$$

so that applying [12, Theorem 6.3.2] yields the claim (for example $\alpha=1/16$ suffices). We closely follow the proof of [12, Theorem 6.3.9] to construct $f_{0,m}$, and use ideas inspired by [2] to choose the remaining parameters of θ_m.

Define

$$f_{0,0}=1,\quad f_{1,0}=f_{0,0}+\zeta\psi_{2,0},$$
$$\psi_{2,0}(x)=\sqrt{3}(2x-1).$$

Note that $f_{0,0},f_{1,0}\geq3/4$ pointwise (recall we assumed $\zeta\leq(4\sqrt{3})^{-1}$) and hence any small perturbations of these will remain bounded away from zero.

We choose perturbations $f_{0,m}$ of f_0 to satisfy the second condition of equation (29), and we choose the remaining parameters $f_{1,m}$, p_m,q_m to ensure the Kullback–Leibler condition holds. Proposition 12, which upper bounds the KL divergence by a " distance" ρ will be of help for the latter.

Define the parameters $\theta_m = (p_m, q_m, f_{0,m}, f_{1,m})$ as follows: First, choose $\phi_{1,m} = -1 + c\delta$ and $\phi_{2,m} = \epsilon$ for all $m \geq 0$ and define p_m, q_m according to the inversion formulae in Lemma 1. Next, for $m \geq 1$, for g_m to be chosen define

$$f_{0,m} = f_{0,0} + g_m, \quad f_{1,m} = f_{1,0} - \frac{1+\phi_1}{1-\phi_1} g_m.$$

Writing $\psi_{1,m}, \psi_{2,m}, \phi_{3,m}$ for the corresponding alternative parametrisation as in Section 2.1, the above choice ensures that $\psi_{1,m} = \psi_{1,0}$ regardless of the choice of g_m. We will choose g_m (depending on n) such that $\|\psi_{2,m} - \psi_{2,0}\|_{L^2} \to 0$ (uniformly in m) as $n \to \infty$ so that in particular it is less than 2 eventually, hence

$$\langle \psi_{2,m}, \psi_{2,0} \rangle = 1 - \tfrac{1}{2} \|\psi_{2,m} - \psi_{2,0}\|_{L^2}^2 \geq 0.$$

Under the condition that $\phi_{3,m} \asymp \zeta$, one sees that

$$\rho\big((\phi,\psi)(\theta_m); (\phi,\psi)(\theta_0))\big) = C \max\Big\{ \delta\epsilon\zeta |\phi_{3,m} - \phi_{3,0}|, \delta\epsilon\zeta^2 \|\psi_{2,m} - \psi_{2,0}\|_{L^2} \Big\}.$$

We calculate $f_{0,m} - f_{1,m} = f_{0,0} - f_{1,0} + \frac{2}{2-c\delta} g_m$ and hence, using that $\|f_{0,0} - f_{1,0}\|_{L^2} = \phi_{3,0} = \zeta$,

$$|\phi_{3,m} - \phi_{3,0}| = \|f_{0,m} - f_{1,m}\|_{L^2} - \|f_{0,0} - f_{1,0}\|_{L^2} \leq \tfrac{2}{2-c\delta} \|g_m\|_{L^2},$$

and

$$\begin{aligned} \|\psi_{2,m} - \psi_{2,0}\|_{L^2} &= \Big\| \frac{f_{0,m} - f_{1,m}}{\phi_{3,m}} - \frac{f_{0,0} - f_{1,0}}{\phi_{3,0}} \Big\|_{L^2} \\ &\leq \frac{|\phi_{3,0} - \phi_{3,m}|}{\phi_{3,m}} + \frac{2\|g_m\|_{L^2}}{2 - c\delta\phi_{3,m}} \lesssim \zeta^{-1} \|g_m\|_{L^2}, \end{aligned}$$

yielding

$$\rho\big((\phi,\psi)(\theta_m); (\phi,\psi)(\theta_0))\big) \leq C' \delta\epsilon\zeta \|g_m\|_{L^2}. \tag{30}$$

[provided $c\delta \leq 1$, say, and the condition $\phi_{3,m} \asymp \zeta$ reduces to $\|g_m\|_{L^2} \leq \zeta/3$, say.].

Now we verify that there are M valid choices of g_m such that $f_{0,m}$ and $f_{0,m'}$ are suitably separated in L^2 distance but suitably close in Kullback–Leibler divergence as in (29), and $f_{0,m}$ and $f_{1,m}$ are in the appropriate Sobolev balls. Fix $S \geq s_0$, and let φ_{jk}, $k \leq 2^j$ be a collection of wavelet functions supported in the interior of $[0,1]$ given as scaled translates $\varphi_{jk} = 2^{j/2} \varphi(2^j(\cdot) - k)$ of an S-regular Daubechies wavelet function φ supported in $[1, 2N]$ for some $N = N(S)$. We may choose a collection of $c_0 2^j$ of these functions whose supports are pairwise disjoint for some $c_0 = c_0(S) > 0$; we denote these $\{\varphi_{jp} : 1 \leq p \leq c_0 2^j\}$ in a slight abuse of notation. By the Varsharmov–Gilbert bound [12, Example 3.1.4] there exist $c_1, c_2 > 0$ such that we may choose a set $\mathcal{M} = \{\beta_{m,\cdot} \in \{-1,1\}^{c_0 2^j} : m \leq 2^{c_1 2^j}\}$ for which

$$\sum_p |\beta_{mp} - \beta_{m'p}|^2 \geq c_2 2^j, \quad \forall p' \neq p.$$

Set $g_m = \alpha_1 \sum_p \beta_{m,p} \varphi_{jp}$ for α_1 to be chosen and observe that

$$\|f_{0,m}\|_{B^{s_0}_{2,\infty}} \leq 1 + \|g_m\|_{B^{s_0}_{2,\infty}} = 1 + \alpha_1 2^{js_0} \Big(\sum_p \beta_{m,p}^2 \Big)^{1/2} = 1 + c_0 \alpha_1 2^{j(s_0 + 1/2)},$$

$$\|g_m\|_{L^2}^2 = \alpha_1^2 \sum_p \beta_{m,p}^2 \|\varphi_{jp}\|_{L^2}^2 = c_0 \alpha_1^2 2^j,$$

$$\|f_{0,m} - f_{0,m'}\|^2 = \|g_m - g_{m'}\|_{L^2}^2 = \alpha_1^2 \sum_p |\beta_{m,p} - \beta_{m,p'}|^2 \geq c_2 \alpha_1^2 2^j.$$

The first line ensures that $\|f_{0,m}\|_{B^{s_0}_{2,\infty}} \leq R$ if $\alpha_1^2 \asymp 2^{-j(2s_0+1)}$; note also that consequently $\|f_{1,m}\|_{B^{s_1}_{2,\infty}} \leq 1 + \delta\|g_m\|_{B^{s_1}_{2,\infty}} \lesssim 1 + \delta 2^{j[s_1-s_0]}$. For this choice of α_1, the second line, in conjunction with (30) and Proposition 12 yields that $K(P_m^{(n)}, P_0^{(n)}) \lesssim n\delta^2\epsilon^2\zeta^2 2^{-2js_0}$, so that choosing j such that $2^{j(2s_0+1)} \asymp n\delta^2\epsilon^2\zeta^2$ gives the required bound on Kullback–Leibler divergences in (29). Note also that $\|g\|_\infty \asymp \alpha_1 2^{j/2}$ so that for this choice of j we have $f_{0,m} \geq 1/2, f_{1,m} \geq 1/2$ on $[0,1]$ for n large, hence Proposition 12 indeed applies, and as soon as $(n\delta^2\epsilon^2\zeta^2)^{-s_0/(1+2s_0)} \lesssim \zeta$ we get as needed $\phi_{3,m} \asymp \zeta$. Also, $f_{1,m}$ is in the appropriate Sobolev ball if $\delta^{2s_1+1}(n\epsilon^2\zeta^2)^{s_1-s_0} \lesssim 1$. Finally, for these choices of α_1 and j, the third line yields $\|f_{0,m} - f_{0,m'}\|_{L^2} \gtrsim (n\delta^2\epsilon^2\zeta^2)^{-s_0/(2s_0+1)}$.

We finally prove the general lower bound

$$R_{\text{rough}} = (n\delta^2)^{-s_0/(2s_0+1)},$$

again using a reduction to multiple testing. As before choose $\phi_{1,m} = -1 + c\delta, \phi_{2,m} = \epsilon$, and choose $f_{0,0}, f_{1,0}$ as in proving R_{smooth}. Now set

$$f_{0,m} = f_{0,0} + g_m, \quad f_{1,m} = f_{1,0}.$$

We now have $f_{0,m} - f_{1,m} = f_{0,0} - f_{1,0} + g_m$ which is of the same form as before up to the coefficient $2/(2-c\delta) \in [1,2]$ which no longer appears. The calculations for ρ then go through fundamentally unchanged except that we no longer have $\psi_{1,m} = \psi_{1,0}$, hence

$$\rho\big((\phi,\psi)(\theta_m);(\phi,\psi)(\theta_0)\big) \leq C' \max\big(\delta\epsilon\zeta\|g_m\|_{L^2}, \|\psi_{1,m} - \psi_{1,0}\|_{L^2}\big).$$

We calculate

$$\psi_{1,m} - \psi_{1,0} = \tfrac{1}{2}(1+\phi_{1,m})f_{0,m} + \tfrac{1}{2}(1-\phi_{1,m})f_{1,m} = \tfrac{1}{2}c\delta g_m,$$

hence calculating the upper bound $C''\delta\|g_m\|_{L^2}$ for ρ.

Choosing $M = \lfloor 2^{c2^j} \rfloor$ functions g_m as before, we again choose the factor α_1 proportional $2^{-j(2s_0+1)}$ to ensure $\|f_{0,m}\|_{B^{s_0}_{2,\infty}} \leq R$; note now that $\|f_{1,m}\|_{B^{s_1}_{2,\infty}} = \|f_{1,0}\|_{B^{s_1}_{2,\infty}}$ for all m so that these are suitably bounded.

Where before we chose $2^{j(2s_0+1)} \asymp n\delta^2\epsilon^2\zeta^2$ to obtain the required bound on the KL divergences in equation (29), we now must choose $2^{j(2s_0+1)} \asymp n\delta^2$. This leads to $\|f_{0,m} - f_{0,m'}\|_{L^2} \gtrsim (n\delta^2)^{-s_0/(2s_0+1)}$ so that equation (29) holds with $R_{\text{rough}} = (n\delta^2)^{-s_0/(2s_0+1)}$ in place of R_{smooth}. This yields the claim.