Finite sample analysis of profile M-estimation in the Single Index model

Andreas Andresen *

Weierstrass-Institute,
Mohrenstr. 39,
10117 Berlin, Germany
andresen@wias-berlin.de

Abstract

We apply the results of Andresen and Spokoiny (2014) to analyse a sieve profile quasi maximum likelihood estimator in the single index model with linear index function. The link function is approximated with $C^2$-Daubechies-wavelets with compact support. We derive a Wilks-type phenomenon and Fisher Theorem in a finite sample setup. The approach is based on showing that the conditions of Andresen and Spokoiny (2014) can be satisfied under a set of mild regularity and moment conditions on the index function, the regressors and additive noise. This allows to construct non-asymptotic confidence sets and to derive asymptotic bounds for the estimator as corollaries.

AMS 2000 Subject Classification: Primary 62F10. Secondary 62H12, 62F25
Keywords: semiparametric estimation, single index model, profile estimator

*The author is supported by Research Units 1735 "Structural Inference in Statistics: Adaptation and Efficiency"
1 Introduction

The single-index model supposes that observations \((Y_i, X_i) \in \mathbb{R} \times \mathbb{R}^m\) with \(m \in \mathbb{N}\) satisfy with two functions \(f: \mathbb{R} \to \mathbb{R}\) and \(h: \mathbb{R}^m \to \mathbb{R}\) and with errors \((\varepsilon_i) \in \mathbb{R}\)

\[ Y_i = f(h(X_i)) + \varepsilon_i, \quad i = 1, \ldots, n. \]

It is assumed that the index function \(h\) is known up to some parameter \(\theta \in \mathbb{R}^p\), the so-called single index, such that one writes \(h(\theta, x)\). The single-index model is widely applied in statistics. For example in econometric studies it serves as a compromise between too restrictive parametric models and flexible but hardly estimable purely nonparametric models. Usually the statistical inference focuses on estimating the single index in this model. A lot of research has already been done in this field. For instance Delecroix et al. (1997) show the asymptotic efficiency of the general semiparametric maximum-likelihood estimator for particular examples and in Härdle et al. (1993) the right choice of bandwidth for the nonparametric estimation of the link function is analysed. See McAleer and da Veiga (2008) for a more recent paper that uses a single index model for the covariance matrix of a portfolio to forecast the value at risk. In this paper we will analyse a special case. We assume that with i.i.d. errors \((\varepsilon_i)\)

\[ Y_i = f(X_i \top \theta^*) + \varepsilon_i, \quad i = 1, \ldots, n, \quad (1.1) \]

for some \(f: \mathbb{R} \to \mathbb{R}\) and \(\theta^* \in S_1^{p+} \subset \mathbb{R}^p\), i.i.d errors \(\varepsilon_i \in \mathbb{R}\) with \(\mathbb{E}[\varepsilon_i] = 0\) and \(\text{Var}(\varepsilon_i) = \sigma^2\) and i.i.d random variables \(X_i \in \mathbb{R}^p\) with distribution denoted by \(\mathbb{P}^{X}\). To ensure identifyability of \(\theta^* \in \mathbb{R}^p\) we assume that it lies in the half sphere \(S_1^{p+} := \{\theta \in \mathbb{R}^p : \|\theta\| = 1, \theta_1 > 0\} \subset \mathbb{R}^p\). We assume that the support of the \(X_i \in \mathbb{R}^p\) is contained in the ball of radius \(s_X > 0\). Further we assume that \(f \in \{f : [-s_X, s_X] \to \mathbb{R}\}\) can be well approximated by a variant of orthonormal \(C^3\)-Daubechies-wavelets \((e_k)_{k \in \mathbb{N}}\) (see section 2.1 for details).

For each \(m \geq 1\) let \(f_\eta\) denote a \(m\)-dimensional approximation of \(f \in \mathcal{X}\):

\[ f_\eta \stackrel{\text{def}}{=} \sum_{k=0}^{m} \eta_k e_k, \]

with properly selected coefficients \(\eta = (\eta_1, \ldots, \eta_m)^\top \in \mathbb{R}^m\). Our aim is to analyse the properties of the estimator \(\hat{\theta}_m := \arg\max_{(\theta, \eta) \in \mathcal{T} \mathcal{L}_m(\theta, \eta)}\), where \(m = (j_m - 1)12 + 2^{j_m - 1} + 2^{j_m - 1} + 11 = j_m 12 + 2^{j_m} - 1\) for some \(j_m \in \mathbb{N}\) and where

\[ \mathcal{L}_m(\theta, \eta) = \sum_{i=1}^{n} \|Y_i - f_\eta(X_i \top \theta)\|^2/2, \quad (1.2) \]
and where \( \mathcal{Y} = S_{1}^{p,+} \times \mathbb{R}^{m} \). Ichimura (1993) analysed a very similar estimator in a more general setting based on a "leave one out" kernel estimation of \( \mathbb{E}[Y_{i}|X_{i}^{\top}\theta] \) instead of using \( f_{\eta}(X_{i}^{\top}\theta) \). He shows \( \sqrt{n} \)-consistency and asymptotic normality of his proposed estimator.

We want to illustrate the results from Andresen and Spokoiny (2014) in this setting, since the proposed estimator is a quasi profile maximum likelihood estimator. To have accordance of notation define for \( \upsilon = (\theta, \eta) \)

\[
\mathcal{L}(\upsilon) = \max_{\upsilon \in \mathcal{Y}: \Pi_{\upsilon} = \theta} \mathcal{L}(\upsilon).
\]

Then the profile MLE can be defined as the point of maximum of \( \mathcal{L}(\theta) \):

\[
\tilde{\theta} = \arg\max_{\theta \in \Theta} \tilde{L}(\theta) = \arg\max_{\theta \in \Theta} \max_{\upsilon \in \mathcal{Y}: \Pi_{\upsilon} = \theta} \mathcal{L}(\upsilon),
\]

and we define the semiparametric excess

\[
\mathcal{L}(\tilde{\theta}) - \mathcal{L}(\theta^{*}) = \max_{\upsilon \in \mathcal{Y}} \mathcal{L}(\upsilon) - \max_{\upsilon \in \mathcal{Y}: \Pi_{\upsilon} = \theta^{*}} \mathcal{L}(\upsilon).
\]

In Andresen and Spokoiny (2014) the results are derived even under miss specification, that is in the case that the true data distribution \( \Pi^{P} \) does not belong to the considered family \( (\Pi_{\upsilon}, \upsilon \in \mathcal{Y}) \). Equivalently, one can say that \( \mathcal{L}(\upsilon) \) is the quasi log-likelihood function on \( \mathcal{Y} \). The "target" value \( \upsilon^{*} = (\theta^{*}, \eta^{*}) \) of the parameter \( \upsilon \) can defined by

\[
\upsilon^{*} = \arg\max_{\upsilon \in \mathcal{Y}} \mathbb{E}\mathcal{L}(\upsilon).
\]

The key result of Andresen and Spokoiny (2014) claims that under the set of conditions of Section B one gets that on a set \( \Omega(x) \) the semiparametric excess can be approximated by a quadratic form \( \|\hat{\xi}\|^{2}/2 \)

\[
\tilde{L}(\theta) - \tilde{L}(\theta^{*}) \approx \|\hat{\xi}\|^{2}/2, \quad (1.3)
\]

while the profile MLE estimator \( \tilde{\theta} \) can be expanded in the form

\[
\hat{D}(\tilde{\theta} - \theta^{*}) \approx \hat{\xi}.
\]

This result is accompanied with the bound

\[
\Pi(\Omega(x)) \geq 1 - C e^{-x},
\]

where \( C > 0 \) is a fixed constant. We call the symmetric matrix \( \hat{D}^{2} \in \mathbb{R}^{p \times p} \) influence matrix and - in the case of correct specification - it is equal to the covariance of the
efficient influence function in the known semiparametric approach; see Kosorok (2005). The random vector \( \tilde{\xi} \in \mathbb{R}^p \) satisfies \( E\tilde{\xi} = 0 \) and \( E\|\tilde{\xi}\|^2 \cong p \). The deviation properties of \( \|\tilde{\xi}\|^2 \) resemble those of a chi-squared distributed random variable with \( p \) degrees of freedom just as in the Wilks phenomenon. From (1.3) one obtains a number of important and informative corollaries such as the construction of confidence and concentration sets. The message is, once conditions on the model are found that ensure that the set of conditions from Appendix A in Andresen and Spokoiny (2014) are satisfied, the results apply and confidence and concentration results follow as Corollaries.

In Section 2 we will summarize the conditions necessary to ensure that the results of Andresen and Spokoiny (2014) apply. In Section 3 we collect the technical calculations and arguments that provide the main results.

2 Main results

Under the same assumptions on \( f \) as in the introduction we can give a list of smoothness conditions on the function \( f \in X \) and on the basis used for the approximation and concentration assumptions on the errors to satisfy the conditions of Section B.

2.1 Choice of basis

To control the bias of the sieve estimator \( \tilde{\theta}_m \in \mathbb{R}^p \) with the proposed approach we can not use any basis \( (e_k)_{k \in \mathbb{N}} \) in \( L^2([-s_X, s_X]) \). We need to show in the proof of Lemma A.2 that the following terms vanish as \( m \to \infty \)

\[
\int_{\mathbb{R}} e_{m+k}(x)e_{m+l}(x)d_{X^\top \theta^*}(x)dx; \ l, k \in \mathbb{N},
\]

(2.1)

where \( d_{X^\top \theta^*} \) denotes the density of \( X^\top \theta^* \in \mathbb{R} \). But terms as in (2.1) do not vanish just for any Fourier basis of \( L^2([-s_X, s_X]) \). It turns out that an orthonormal wavelet basis is suitable for this purpose. For high indeces \( k \in \mathbb{N} \) the support of each wavelet \( e_k \) is contained in a small interval on which the density \( f_{X^\top \theta^*} \) can be well approximated by a constant. Due to orthogonality and shrinking supports of the basis the term in (2.1) can be shown to diminish sufficiently fast for a Lipschitz continuous density \( f_{X^\top \theta^*} \) (see Lemma A.2). The trouble is that our approach relies on smoothness of the basis elements. Consequently we need a smooth orthogonal wavelet basis on an interval. Thanks to Daubechies (1992) such a basis \( (\Psi_{n,m})_{n,m \in \mathbb{Z}} \) is available on \( L^2(\mathbb{R}) \). We can use this basis to construct an appropriate basis \( (e_k)_{k \in \mathbb{N}} \) on the interval. This basis will have all the properties needed for the proof of Lemma A.2 and thus will allow us to control the bias.
To understand the choice of this basis \((e_k)_{k \in \mathbb{N}}\) we first have to briefly explain how the Daubechies wavelets are derived. To ease understanding we adopt the notation of that paper. Starting with a scaling function \(\phi : \mathbb{R} \to \mathbb{R}\) where \(\|\phi\|_{L^2(\mathbb{R})} = 1\) one obtains a sequence of nested spaces

\[ V_j = \text{span}\{2^{-j/2}\phi(2^{-j} \cdot -n); n \in \mathbb{Z}\} \subset L^2(\mathbb{R}), \ldots \subset V_1 \subset V_0 \subset V_{-1} \subset \ldots \subset L^2(\mathbb{R}). \]

If the scaling function \(\phi : \mathbb{R} \to \mathbb{R}\) satisfies certain properties one can show that \(\bigcup_{n \in \mathbb{Z}} V_n = L^2(\mathbb{R})\) and that \((2^{-j/2}\phi(2^{-j} \cdot -n))_{m \in \mathbb{Z}}\) is an orthonormal basis in \(V_j \subset L^2(\mathbb{R})\) for every \(j \in \mathbb{Z}\). Denote for each \(j \in \mathbb{Z}\) by \(W_j \subset L^2(\mathbb{R})\) the orthogonal complement of \(V_{j+1} \subset L^2(\mathbb{R})\) in \(V_j \subset L^2(\mathbb{R})\). This gives

\[ V_j = V_{j+1} \oplus W_{j+1} = \bigoplus_{k > j, k \in \mathbb{Z}} W_k, \text{ such that } L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j. \tag{2.2} \]

The idea of Daubechies wavelets is to find a function \(\psi \in W_1\) that satisfies with \(\psi_{j,n} \overset{\text{def}}{=} 2^{-j/2}\psi(2^{-j} \cdot +n)\)

\[ W_j = \text{span}(\psi_{j,n}; n \in \mathbb{Z}), \quad \langle \psi_{j,n}, \psi_{j,n'} \rangle_{L^2} = \delta_{n,n'}, n, n' \in \mathbb{Z}. \]

This is indeed possible. For this denote

\[ h_n = \langle \phi, \phi(2 \cdot +m) \rangle, m \in \mathbb{Z}, \text{ i.e. } \phi = \sqrt{2} \sum_{m \in \mathbb{Z}} h_m \phi(2 \cdot -m), \]

and define

\[ \psi = \sqrt{2} \sum_{m \in \mathbb{Z}} (-1)^{n-1}h_{-n-1}\phi(2 \cdot -n). \]

Theorem 6.3.6, Lemma 6.2.2 and the table at the end of Section 7.3.1 of Daubechies (1992) tell us that there exists a scaling function \(\phi_7 : \mathbb{R} \to \mathbb{R}\) for which the associated family \(\psi_{j,n} \overset{\text{def}}{=} 2^{-j/2}\psi(2^{-j} \cdot +n)\) satisfies

\[ (\psi_{j,n})_{j,n \in \mathbb{Z}} \text{ ONB of } L^2(\mathbb{R}), \text{ support}(\psi) \subset [0,13], \psi \in C^2(\mathbb{R}). \tag{2.3} \]

So we obtain a well suited basis for \(L^2(\mathbb{R})\) but we only need one for \(L^2([-sX,sX])\). We could simply embed

\[ L^2([-sX,sX]) \to L^2(\mathbb{R}), \quad f(\cdot) \mapsto f(\cdot)1_{[-sX,sX]}, \]

and use that basis but this would mean that we have to include basis functions \(\psi_{j,n} \in L^2(\mathbb{R})\) for positive \(j \in \mathbb{N}\) as well. We want to avoid this as it is not necessary for our
Finite sample analysis of profile M-estimation in the Single Index model

purpose. Instead we do the following: First adapt the scale and support of the basis and the corresponding shift operation to the interval via redefining

\[ \phi_{7,sX}(t) = (2sX)^{-1/2}\phi_7((2sX)^{-1}t + 1), \psi_{sX}(t) = (2sX)^{-1/2}\psi((2sX)^{-1}t + 1) \]

The associated wavelet basis \( \psi_{j,n} \) still satisfies all properties from (2.3) where the support is adapted to read \([-sX, 25sX]\). Next note that (2.2) and the definition of the sub spaces implies

\[ L^2(\mathbb{R}) = V_0 \oplus \bigoplus_{j \in \mathbb{N}} W_{-j}, \]

where the definition is adapted to read \( V_j = \text{span}\{2^{-j/2}\phi_{7,sX}(2^{-j} \cdot -nsX); n \in \mathbb{Z}\} \subset L^2(\mathbb{R}) \). As we only have to approximate functions that are nonzero on \([-sX, sX]\) this suggest the following basis: for \( k = j_k12 + 2^{j_k} + r_k \in \mathbb{N} \) where \( j_k \in \mathbb{N}_0 \) and \( r_k \in \{0, \ldots, 2^{j_k} + 11\} \) we set

\[ e_k \overset{\text{def}}{=} \begin{cases} \phi_{7,sX}(t + (k - 1)sX) & \text{if } k \leq 12, \\ \psi_{-j_k,r_k-12} & \text{if } k > 12. \end{cases} \]

So in words we include all elements of a basis for \( L^2(\mathbb{R}) \) which have a support with nonempty intersection with \([-sX, sX]\). Restricting the pre image of the elements of the closed span of these functions to \([-sX, sX]\) we end up with a basis for \( L^2([-sX, sX]) \), that is contained in \( C^2(\mathbb{R}) \) and satisfies for any \( l, k \in \mathbb{N} \) with \( k = j_l12 + 2^{j_l} + r_l \in \mathbb{N} \)

\[ \langle e_l, e_k \rangle_{L^2(\mathbb{R})} = \delta_{l,k}, \quad |\text{support}(\psi_k)| \leq 2^{j_k}26sX. \]

Further this basis has another useful property that will come in handy in the proof of Lemma A.2: For any \( k \in \mathbb{N} \) with \( k = j_k12 + 2^{-j_k} + r_k \in \mathbb{N} \) it holds

\[ \left| \left\{ l = j_l12 + 2^{j_l} + r_l \middle| r_l \in \{0, \ldots, 2^{j_l} + 11\}, \text{supp}(e_k) \cap \text{supp}(e_l) \neq \emptyset \right\} \right| \leq 2^{(j_l-j_k)13}. \quad (2.4) \]

### 2.2 Assumptions

To apply the technique presented in Andresen and Spokoiny (2014) we need a list of assumptions. We denote this list of conditions by \( (A) \). We start with conditions on the regressors \( X \in \mathbb{R}^p \):

\( (\text{Cond}_X) \) The measure \( \mathbb{P}^X \) is absolutely continuous with respect to the Lebesgue measure. The Lebeque density \( f_X \) of \( \mathbb{P}^X \) is only positive on the ball \( B_{sX}(0) \subset \mathbb{R}^p \) and
Lipschitz continuous with Lipschitz constant $L_{f_X} > 0$. Further we assume that for any $\theta \perp \theta^*$ with $||\theta|| = 1$ we have $\text{Var} \left( X^\top \theta \big| X^\top \theta^* \right) > \sigma_{X|\theta^*}^2$ for some constant $\sigma_{X|\theta^*}^2 > 0$ that does not depend on $X^\top \theta^* \in \mathbb{R}$. Also assume that $f_X > c_{d_X} > 0$ on $B_{s_x}(0) \subset \mathbb{R}^p$.

**Remark 2.1.** We only assume bounded support of the regressors $(X_i) \subset \mathbb{R}^p$ for simplicity. This condition could be relaxed to a qualified probabilistic deviation bound of the kind $\mathbb{P}(||X|| \geq s_X + x) \lesssim e^{-x}$. Further $\text{Var} \left( X^\top \theta^* \big| X^\top \theta^* \right) = 0$ would mean that $X^\top \theta^* = h(X^\top \theta^*)$ for some function $h : \mathbb{R} \to \mathbb{R}$. But then we would have for any $(\alpha, \beta) \in \mathbb{R}^2$ with $\alpha^2 + \beta^2 = 1$ that

$$f(X^\top (\alpha \theta^* + \beta \theta^*)) = f(\alpha X^\top \theta^* + \beta h(X^\top \theta^*)) \overset{\text{def}}{=} g_{\alpha, \beta}(X^\top \theta^*),$$

such that the problem would no longer be identifiable. We bound $f_X > c_{d_X} > 0$ on $B_{s_x}(0) \subset \mathbb{R}^p$ again to ensure identifyability.

Of course we need some regularity of the link function $f \in \{ f : [-s_X, s_X] \mapsto \mathbb{R} \}$:

**($\text{Cond}_f$)** For some $\eta^* \in l^2$

$$f = f_{\eta^*} = \sum_{k=1}^{\infty} \eta_k^* e_k,$$

(2.5)

where with some $\alpha > 2$ and a constant $C_{||\eta^*||} > 0$

$$\sum_{l=0}^{\infty} l^{2\alpha} \eta_l^* \leq C_{||\eta^*||} < \infty.$$  

For the large deviations of the MLE we need the following condition:

**($\text{Cond}_{X\theta^*}$)** It holds true that $\mathbb{P}(|f'_{\eta^*}(X^\top \theta^*)| > c_{f'_{\eta^*}}) > c_{f'_{\eta^*}}$ for some $c_{f'_{\eta^*}}, c_{f'_{\eta^*}} > 0$.

**Remark 2.2.** Note that a condition of this kind is necessary to ensure identifyability. Otherwise the function $f$ would be $\mathbb{P}^X$-almost surely constant. But for a constant function $C(x) \equiv c$ any $\theta \in \mathbb{R}^p$ solves $C(X^\top \theta) = c$.

To be able to apply the finite sample device we need constraints on the moments of the additive noise:

**($\text{Cond}_{\varepsilon}$)** The errors $(\varepsilon_i) \in \mathbb{R}$ are i.i.d. with $\mathbb{E}[\varepsilon_i] = 0$, $\text{Cov}(\varepsilon_i) = \sigma^2$ and satisfy for all $|\mu| \leq \tilde{g}$ for some $\tilde{g} > 0$ and some $\tilde{\nu}_\varepsilon > 0$

$$\log \mathbb{E}[\exp \{ \mu \varepsilon_1 \}] \leq \overline{\nu}_\varepsilon^2 \mu^2/2.$$  

Finally to be able to control the large deviations of the MLE we impose

**($\text{Cond}_Y$)** $Y \subseteq Y_e(\sqrt{m} \nu^2) \subset \mathbb{R}^{p+m}$ with $\nu^2 \in \mathbb{R}$. Such that $d_Y \overset{\text{def}}{=} \text{diam}(Y) < \infty$. 

2.3 Some important objects

In this Subsection we introduce the likelihood functional, the profile estimator and some important objects that are relevant for our results.

For given \( p^* = p + m \), set \( \Pi_{p^*}v = (v_1, \ldots, v_{p^*}) = (\theta, \Pi_{m}\eta) \in \mathbb{R}^{p^*} \). We represent the full parameter \( v \in \mathbb{R}^\infty \) in the form

\[
v = (\theta, f) = (\Pi_{p^*}v, \tau) = (\theta, \Pi_{m}\eta, \tau)
\]

where \( \tau = (\eta_{m+1}, \ldots)^\top \) stands for the remaining components of the expansion (2.5). Then introduce the following functional for statistical inference:

We define the sieve estimator \( \tilde{v}_m \), its possibly biased target \( v^*_m \) and the full oracle \( v^* \in \mathbb{R}^\infty \)

\[
\begin{align*}
\tilde{v}_m &= \arg\max_{v \in \mathbb{R}^{p^*}} \mathcal{L}(v, 0), \\
v^*_m &= (\theta^*_m, \eta^*_m) = \arg\max_{v \in \mathbb{R}^{p^*}} \mathbb{E}[\mathcal{L}(v, 0)], \\
v^* &= (\Pi_{p^*}v^*, \tau^*) = \arg\max_{v \in l^2} \mathbb{E}[\mathcal{L}(v)],
\end{align*}
\]

where \( \mathcal{L}(\cdot) \) is the likelihood functional from (1.2).

Remark 2.3. We will see that \( (v^*_m, 0) \in l^2 \) lies close the true point \( v^* \in l^2 \) but we will not proof that it is unique. In the following we will denote by \( v^*_m \) the set of maximizers. We neither proof or use uniqueness of the pMLE either. In the following we will always make statements about \( \tilde{\theta}_m \in \mathbb{R}^p \), whereby we mean any element of the set of maximizers of the profiled likelihood functional. Non-uniqueness is not a problem, as the concentration on the local set \( Y_0 \) is ensured via an upper function approach. For details see Spokoiny (2012) Corollary 4.4.

We define the information operator \( \mathcal{D}^2 \) similarly to the Fisher information matrix as the Hessian operator of the expected value of the likelihood functional:

\[
\mathcal{D}^2 \overset{\text{def}}{=} -\nabla^2 \mathbb{E}\mathcal{L}(v^*) = -\nabla^2 \mathbb{E}\mathcal{L}(\theta^*, f^*).
\]

Consider the following block representations of of the information operator:

\[
\mathcal{D}^2 = \begin{pmatrix}
D^2 & A_{\theta, \tau} \\
A_{\tau, \theta}^\top & \mathcal{H}_{\tau, \tau}^2
\end{pmatrix} = \begin{pmatrix}
D^2 & A_{m} & A_{\theta, \tau} \\
A_{m} & \mathcal{H}_{m}^2 & A_{\eta, \tau} \\
A_{\tau, \theta} & A_{\tau, \eta} & \mathcal{H}_{\tau, \tau}^2
\end{pmatrix}.
\]
where \( A_{\nu, \nu} \) is an operator from \( \mathbb{R}^\infty \) to \( \mathbb{R}^{p+m} \). Define the smallest eigenvalue \( c_0 \) such that

\[
\lambda_{\min}(D)/\sqrt{n} > 0.
\]

In Lemma A.4 we derive that

\[
c_0 \geq \lambda_{\min}(H)/4,
\]

Further we introduce the influence matrix and the score

\[
\bar{D}_m = \Pi_\theta D_m^{-2} \Pi_\theta^T, \quad \bar{\xi}_m = \nabla_\theta \mathcal{L} - A_m H_m^{-2} \nabla_\eta \mathcal{L}.
\]

We have

\[
\sigma^2 \bar{D}_m^{-2} = \text{Cov} (\bar{\xi}_m) \quad \text{where we remind that} \quad \sigma^2 = \text{Var}(\varepsilon_i) \quad \text{is the variance of the additive errors in (1.1)}.
\]

For certain constants we need the following symbols: \( L_{\nabla \theta} \) denotes the Lipshitz constant of the gradient of the canonical parametrization of the sphere in \( \mathbb{R}^p \) and \( d_{X^\top \theta^*} \in \{ f : [-sX, sX] \to \mathbb{R} \} \) the density of \( X^\top \theta^* \in \mathbb{R}^p \). We also need \( C_{\nabla_\nu} \) such that

\[
\gamma_2^m (\nu) = \text{Cov} \left( \nabla_\nu \mathcal{L} (\nu) \right).
\]

### 2.4 Properties of the Wavelet Sieve profile MLE (WSpMLE)

This sections presents the main results of this analysis. We suppose that a sufficiently large constant \( x > 0 \) is fixed. It appears that our results are slightly different in two zones separated by the value \( x_0 \). To keep results shorter we only show the case \( x \leq x_0 \) as it is done in Andresen and Spokoiny (2014). See Section C in that paper for more details.

We get the following result by applying Theorem 2.1 of Andresen and Spokoiny (2014)

**Proposition 2.1.** Assume \((A)\), that \( m^{-2(\alpha+1)} n \to 0 \) and that \( p^*/\sqrt{n} \to 0 \). If \( n \in \mathbb{N} \) is large enough, it holds with probability greater than 

\[
1 - 6e^{-x} - \exp \left\{ -n/m^5 \right\} - \exp \left\{ -nc(Q)/4 \right\}
\]

\[
\| 2 \bar{L}(\tilde{\theta}_m, \theta^*_m) - \| \bar{\xi}_m \| \| \leq \tau_p \left( p^* + x \right)^5/2 / \sqrt{n} + c_0 \left( p^* + x \right)^5/2 / n,
\]

\[
\| \bar{D}_m (\tilde{\theta}_m - \theta^*_m) - \bar{\xi} \| \leq c_0 \left( p^* + x \right)^5/2 / \sqrt{n}
\]

where \( c_0 > 0 \) and

\[
\tau_p \leq \left( 2 + \frac{1 + \rho}{1 - \rho} \right) c_0 \left( p^* + x \right)^5/2 / \sqrt{n} + \frac{1 + \rho}{1 - \rho} \sigma \sqrt{6 \sqrt{p} + x}.
\]
Remark 2.4. The constant \( C_\phi > 0 \) is a polynomial of \( \| \psi \|_\infty, \| \psi' \|_\infty, \| \psi'' \|_\infty, C_\| F \|, L\nabla \phi, s_X, c_{D^1}, c_{\psi_T}, \sigma \) and \( \| f_X \theta \|_\infty \) that is independent of \( x, n, p^* \). The constant \( c(Q) > 0 \) is derived in the proof of Lemma A.12 and also does not depend on \( x, n, p^* \).

Remark 2.5. The necessary size of \( n \in \mathbb{N} \) is determined by the ratio \( p^{5/2}/\sqrt{n} \rightarrow 0 \). In the proof of Lemma 3.6 we impose conditions on \( n \in \mathbb{N} \) of the kind

\[
p^{5/2}/\sqrt{n} \leq C_1^{-1}, \quad m^{-2\alpha-1}n \leq C_2^{-1},
\]

for certain constants \( C_1, C_2 > 0 \) that are polynomials of \( \| \psi \|_\infty, \| \psi' \|_\infty, \| \psi'' \|_\infty, C_\| F \|, L\nabla \phi, s_X \).

As is explained in Andresen and Spokoiny (2014) these two results imply all other relevant results for statistical inference. But so far we only addressed the behavior of the sieve pMLE with respect to the possibly biased target \( \theta^*_m \in \mathbb{R}^p \) and with a weighting matrix that depends on the dimensions \( m \in \mathbb{N} \) of the nuisance parameter \( \eta \in \mathbb{R}^m \). The next result will specify the finite sample properties of \( \hat{D}(\theta - \theta^*) \in \mathbb{R}^p \) where

\[
\hat{D}^{-2} = \Pi_\theta D^{-2} \Pi_\theta^\top \in \mathbb{R}^{p \times p},
\]

which is up to \( \sigma^2 = \text{Var}(\varepsilon) > 0 \) the lower bound for the variance of regular estimators of \( \theta^* \in \mathbb{R}^m \), due to the general results on efficiency in semiparametric models (van der Vaart and Wellner (1996), Theorem 3.11.2 p. 414, setting \( \nu(\mathbb{P}_n) = \theta \)). We get the following result.

Proposition 2.2. Assume (A). If \( m^{-2\alpha+1}n \rightarrow 0 \), \( p^{5/2}/\sqrt{n} \rightarrow 0 \) and if \( n \in \mathbb{N} \) is large enough it holds with probability greater \( 1 - 6e^{-x} \)

\[
\| \hat{D}(\tilde{\theta}_m - \theta^*) - \hat{\xi}_m(\nu_m^*) \| \leq \left( 2 + \sqrt{\frac{1 + \rho^2}{1 - \rho^2}} C_\Delta \frac{(p^* + x)^{5/2}}{\sqrt{n}} \right) + \alpha(m),
\]

\[
|2 \hat{L}(\theta^*_m, \theta^*) - \| \hat{\xi}_m(\nu_m^*) \|^2 | \leq 2 \left( \rho_p + C_\Delta \frac{(p^* + x)^{5/2}}{\sqrt{n}} \right) \frac{(p^* + x)^{5/2}}{\sqrt{n}} + \left( \frac{\delta_p^* + C_\Delta (p^* + x)^{5/2}}{\sqrt{n}} \right)^2 + 2\alpha(m)^2 + 2\alpha(m)\sqrt{6\sigma\sqrt{p + x}},
\]

where

\[
\alpha(m) \leq \sqrt{\frac{1 + \rho^2}{1 - \rho^2}} C_1 \sqrt{mn}^{-\alpha - 1/2},
\]

and

\[
\delta_p^* \overset{\text{def}}{=} \left( 2 + \frac{1 + \rho}{1 - \rho} \right) C_\Delta \frac{(p^* + x)^{5/2}}{\sqrt{n}} + \alpha(m).
\]
Further as $n \to \infty$

$$
\| \hat{D}(\theta_m - \theta^*) - \hat{\xi}_m(\upsilon_m^*) \| \overset{P}{\to} 0, \\
\hat{D}(\theta_m - \theta^*) \overset{w}{\to} N(0, \sigma^2 I_p), \\
2\hat{L}(\theta_m, \theta^*) \overset{w}{\to} \chi^2_p.
$$

**Remark 2.6.** The constraints $m^{-(2\alpha+1)} n \to 0$ and $p^*^{5/2}/\sqrt{n} \to 0$ are exclusive for $\alpha \leq 2$. But note that if $0 < \alpha - 2 =: \epsilon$ and $m \geq n^{1/5-\delta}$ with $\delta > 2\epsilon/(25 + 5\epsilon)$ we get

$$
m^{-2\alpha-1} n^1 \leq n^{-(1+2\epsilon/5)+\delta(2\alpha+1)+1} = n^{-2\epsilon/5+\delta(5+2\epsilon)} \to 0,
$$

such that $n = o(m^{2\alpha+1})$ and $p^* = o(n^{1/5})$.

**Remark 2.7.** Remember that

$$
\tilde{L}(\theta) \overset{\text{def}}{=} \max_{\eta \in \mathbb{R}^m} \mathcal{L}(\theta, \eta),
$$

where it is important to note that the maximization is restricted to the finite dimensional space $\mathbb{R}^m$.

### 3 Details

In this section we lay out our strategy. It is our aim to apply Theorem 2.1 of Andresen and Spokoiny (2014). For this we need a list of definitions.

Remember that

$$
\mathcal{V}_m^2 = \text{Cov} \left( \nabla_{p+m} \mathcal{L}(\upsilon^*) \right), \quad \mathcal{D}_m^2 = -\nabla^2_{p+m} \mathbb{E} \mathcal{L}(\upsilon^*),
$$

and $\upsilon_m^* = \text{argmax} \mathbb{E} \mathcal{L}_m(\upsilon)$. Define for some $x, r > 0$ the quantity $\diamond(x, x) > 0$

$$
\diamond(x, x) \overset{\text{def}}{=} \left( \delta(x) + 6\nu_1 \omega \tilde{g}(x, Q) \right) r,
$$

where the constant $\nu_1$ is from condition (ED1) in Section B and where $\tilde{g}(x, Q) \equiv \sqrt{x + p^*}$ for moderate choice of $x > 0$ (see Appendix E of Andresen and Spokoiny (2014)).

The value $r_0 > 0$ is defined by the condition

$$
\mathbb{P}(\tilde{v} \in Y_{c}(r_0)) \geq 1 - ce^{-x},
$$

and for some quadratic matrix $\mathbb{B}$ the quantity $\tilde{g}(x, \mathbb{B})$ is defined in (see Appendix C of Andresen and Spokoiny (2014)).
Also the random variable $\bar{v}_{\theta^*} \in \mathcal{T}$ is defined as the maximizer of $\mathcal{L}(v,v^*)$ subject to $Pv = \theta^*$:

$$\bar{v}_{\theta^*} \overset{\text{def}}{=} \arg\max_{v \in \mathcal{T}} \mathcal{L}(v,v^*) \quad \text{subject to} \quad P_{\theta^*} = \theta^*.$$  

With $B = D^{-1}V^2D^{-1}$ and $\bar{B} = \bar{D}^{-1}\bar{V}^2\bar{D}^{-1}$ Theorem 2.1 of Andresen and Spokoiny (2014) combined with Lemma 2.3 of that paper gives with

$$r_1 \overset{\text{def}}{=} \left( \frac{1}{2}(x,B)/(1 - \rho) + \hat{\delta}(r_0,x) \right) \wedge r_0. \quad (3.1)$$

**Theorem 3.1.** With central point $v^0 = v^* \in \mathbb{R}^p$ and matrices $V_0^2 = V^2$ and $D_0^2 = D^2$ assume that $(ED_0)$, $(ED_1)$, $(L_0)$ and $(I)$ are satisfied; see Section B. Find a radius $r_0 > 0$ such that $P(v,v^* \in \mathcal{Y}(r_0)) \geq 1 - e^{-x}$. Then it holds on a set $\Omega(x) \subseteq \Omega$ with probability $(1 - 6e^{-x})_+ \geq 0$

$$|2\bar{L}(\hat{\theta},\theta^*) - \|\hat{\xi}\|^2| \leq 2\left( r_p + \hat{\delta}(r_1,x) \right) \hat{\delta}(r_1,x), \quad \|\bar{D}(\hat{\theta} - \theta^*) - \hat{\xi}\| \leq \hat{\delta}(r_1,x).$$

where

$$r_p \overset{\text{def}}{=} \left( 2 + \frac{1 + \rho}{1 - \rho} \right) \hat{\delta}(B,r_0,x) + \frac{1 + \rho}{1 - \rho} \hat{\delta}(\bar{B},x).$$

Our strategy will be the following. First we will show that the conditions $(ED_0)$, $(ED_1)$, $(L_0)$, $(I)$, of the previous Theorem can be satisfied under the assumptions $(A)$. Further we will show that the conditions $(E\mathbf{r})$ and $(L\mathbf{r})$ from B are met. This will allow to determine $r_0 > 0$. The subsequent analysis will then serve to determine the necessary size of $n \in \mathbb{N}$ that allows to obtain good bounds for $\hat{\delta}(B,r_0,x) \in \mathbb{R}$.

### 3.1 Conditions satisfied

In this section we show that the conditions of section B are satisfied. First we derive an apriori bound for the distance between the target $v^*_m \in \mathbb{R}^p \times \mathbb{R}^m$ and the true parameter $v^* \in \mathbb{R}^p \times \mathbb{R}^\infty$

**Lemma 3.2.** Assume $(A)$ and set

$$r^* = 16\left( C(m)\|f_X\|_{\infty}C_{\|f\|_{\infty}} + C(m)26\sqrt{28}\sqrt{s_{X}^{p+1}}L_{f_X}\|\psi\|_{\infty}C_{\|f\|_{\infty}}^2 \right)^{1/2} \quad (3.2)$$

$$\sqrt{nm} - (1 + 2\alpha)/\sqrt{m}.$$
If \( n \in \mathbb{N} \) is large enough to ensure that
\[
C_{\delta,\infty} \left( m^{3/2} \vee \frac{r^*m^3}{\sqrt{n}} \right) 2r^* \leq c_D \frac{\sqrt{n}}{c_D \sqrt{n}}.
\]
Then we get \( \|D_m^{1/2}(\psi_m^* - \psi^*)\| \leq r^* \).

**Remark 3.1.** The constant \( C_{\delta,\infty} > 0 \) is a polynomial of \( \|\psi\|_\infty, \|\psi'\|_\infty, \|\psi''\|_\infty, c_{\|f\|}, \)
\( L\nabla\phi, s_X, \|f_X^\top \theta^*\|_\infty \); \( c_D^{-1} \) that is independent of \( x, n, p^* \). The constant \( c_D > 0 \) is the smallest eigenvalue \( c_D \overset{\text{def}}{=} \lambda_{\min}(D)/\sqrt{n} \) and as shown in Lemma A.4 satisfies
\[
c_D \geq \lambda_{\min}(H) c^2_{\|f\|} c_{\|f\|}^2 \sigma^2 \|X\|_2 / (2s_X C_{\|f\|}^\infty) \).\]

Now we show that the general conditions from section B are met under the assumptions (A):  

**Lemma 3.3.** Assume the conditions (A). Then we get with \( \psi^0 = \psi_m^* \in IR^{p*} \) and
\[
\mathcal{V}_0^2 = \text{Cov}(\nabla \mathcal{L}(\psi^*)), \quad \mathcal{D}_0^2 = -\nabla^2 \mathcal{E} \mathcal{L}(\psi^*),
\]
the conditions of section B in the following form:

\((\mathcal{E}D_0)\) with
\[
g = \sqrt{n} \sigma^{-1} c_D \tilde{g}(C_{\|\eta^*\|} + 1) \sqrt{26s_X \|\psi'\|_\infty + \sqrt{13} \|\psi''\|_\infty \sqrt{m}}^{-1},
\]
\[
\nu_m^2 = \tilde{\nu}_2^2 \left( \|\mathcal{V}_m^1(\psi^*)\| m_2^2(\psi_m^*)^2 + 1 \right),
\]
\((\mathcal{E}r)\) with
\[
g(\mathbf{r}) = c_D \tilde{g}(C_{(\mathcal{E}r)}) m^{-3/2} \sqrt{n},
\]
\[
\nu_{r,m}^2 = \tilde{\nu}_2^2 \left( c_{\mathcal{V}_m^0} + c_{\mathcal{V}_m^1} \right) \sqrt{n},
\]
where \( C_{\mathcal{V}_m^0} = \sup_{\psi \in \mathcal{V}_m^0(\sqrt{n})} \|\mathcal{V}_m^2(\psi)^{-1}\mathcal{V}_m^2(\psi^*)\| \) with some \( \mathbf{r}^0 > 0 \) and where one can bound
\[
C_{(\mathcal{E}r)} \leq \sqrt{26} \left( s_X \|\psi'\|_\infty + 3(C_{\|f\|} + 1)\|\psi''\|_\infty s_X + 3\|\psi'\|_\infty s_X + \|\psi'\|_\infty C_{\|\eta^*\|} \sqrt{2L\nabla \phi} \right) \frac{2}{c_D}
\]
\[
+ (C_{\|f\|} + 1) \sqrt{26s_X \|\psi'\|_\infty + \sqrt{13} \|\psi''\|_\infty}.\]
Finite sample analysis of profile M-estimation in the Single Index model

\((\mathcal{ED}_1)\) with

\[ g \overset{\text{def}}{=} \sqrt{n} c_D \rho^{-3/2} C^{-1}(\mathcal{ED}_1), \]

\[ \omega \overset{\text{def}}{=} \frac{2}{\sqrt{n} c_D}, \]

\[ \nu^2_{1,m} = \nu^2_r m^3 C(\mathcal{ED}_1) \left( \frac{1}{c_D^2} + \frac{1}{\sqrt{n}} \right), \]

where

\[ C(\mathcal{ED}_1) = 26 \left( \| s_X \| \psi' \| \infty + 3 \| f \| \| \psi'' \| \infty s_X + 3 \| \psi' \| \infty s_X + \| \psi' \| \infty C \| \eta^*_m \| \sqrt{2L\nabla\Phi} \right). \]

\((\mathcal{L}_0)\) is satisfied where with \( r^* > 0 \) from (3.2)

\[ \delta(r) = \frac{c(\mathcal{L}_0) \left( m^{3/2} \sqrt{\frac{m^3}{\sqrt{n}}} \right) [r + r^*]}{c_D \sqrt{n}}. \]

\((\mathcal{L}_r)\) for \( n \in \mathbb{N} \) large enough with \( b = c(\mathcal{L}_r) > 0 \) and with probability \( 1 - \exp \{-n/m^5\} - \exp \{-nc(\mathcal{Q})/4\} \) as soon as \( r^2 \geq C \rho^* \) for certain constants \( c(\mathcal{Q}), c(\mathcal{L}_r), c > 0 \).

Remark 3.2. The constant \( c(\mathcal{L}_0) > 0 \) is polynomial of \( \| \psi \| \infty, \| \psi' \| \infty, \| \psi'' \| \infty, c \| f \|, L\nabla\Phi, s_X, c_D^{-1}, c_{V^0} \) and \( \| f_X^\top \theta^* \| \infty \) that are independent of \( x, n, p^* \).

Lemma 3.4. Under the assumptions of the last lemma the identifiability conditions from (B.1) and (B.2) are satisfied with \( a^2 = \sigma^2 \) and

\[ \rho^2 \leq 1 - \frac{c_D}{4s_X^2 \| f_X \| \infty C \| \eta^* \|}. \]

Proof. This is clear since \( D \geq \sigma^{-1} V \) that gives (B.1) with \( a^2 = \sigma^2 \) and (B.2) follows from \( V \geq c_D Id \) with Lemma A.17 with

\[ \rho^2 \leq 1 - \frac{ncD}{\lambda_{\max} D_{gg} \wedge \lambda_{\max} D_{ff}} \leq 1 - \frac{c_D}{4s_X^2 \| f_X \| \infty C \| \eta^* \|}. \]

\(\square\)

3.2 Large deviations

Next we determine the necessary size of the radius \( r_0(x) \) defined by

\[ r_0(x) \overset{\text{def}}{=} \inf \{ r > 0 : \mathbb{P}\{ v_m \in \mathcal{T}_0(r) \} \leq 2e^{-x} \}, \]

\[ \mathcal{T}_0(r) \overset{\text{def}}{=} \{ v \in \mathbb{R}^{p^*} : \| D_m (v - v^*_m) \| \leq r \}. \]
Lemma 3.5. We have
\[ r_0 \leq 12 \frac{\bar{D}_r}{c(L,r)} \left( c^2_{D_0} + c^2_{(\xi,r)m^3/\sqrt{n}} \right)^{1/2} \sqrt{x + 4p^*} \]
if
\[ 1 \leq \frac{6c_Dg_{\bar{D}_r}^2 \left( c_{D_0}(r^o)^2 + c^2_{(\xi,r)m^3/\sqrt{n}} \right) \sqrt{n}}{m^{3/2}c(L,r)(1 + \sqrt{x + 4m})} \]

This allows to prove the following lemma:

Lemma 3.6. If \( m^{-2\alpha n} \to 0 \), if \( p^{5/2}/\sqrt{n} \to 0 \) and if \( n \in \mathbb{N} \) is large enough we get
\[ \triangle(r_0) \leq \sqrt{p^* + x}, \]
and
\[ \triangle(r_1, x) \leq c_0 \frac{(x + p^*)^{5/2}}{\sqrt{n}}, \]
where \( c_0 > 0 \) is a polynomial of \( \|\psi\|_{\infty}, \|\psi^0\|_{\infty}, \|\psi''\|_{\infty}, \|f^*\|, L_{\nabla \theta}, s x \).

With these results Proposition 2.1 is merely a corollary of Theorem 3.1 from Andresen and Spokoiny (2014).

3.3 Proof of Proposition 2.2

We prove this claim via showing that the conditions of Corollary 2.3 of Andresen and Spokoiny (2014) are met. To be precise we use the following modified version:

Corollary 3.7. Assume \( \textbf{(bias)} \) and that the conditions \( (E\mathcal{D}_0), (E\mathcal{D}_1) \) and \( (L_0) \) from section B are satisfied for all \( m \geq m_0 \) for some \( m_0 \in \mathbb{N} \) and with \( \mathcal{D}_0 = \nabla_{p+m}^2 \mathcal{E} \mathcal{L}_m(u^*) \in \mathbb{R}^{p^* \times p^*}, \mathcal{V}_0 = \text{Cov} [\nabla_{p+m} \mathcal{L}_m(u^*)] \in \mathbb{R}^{p^* \times p^*} \) and \( \nu^o = \nu^*_m \in \mathbb{R}^{p^*} \). Choose \( r_0(x) > 0 \) such that \( \mathbb{P} (\tilde{\theta}_m, \tilde{\theta}^*_m, \mathcal{V}_0, \mathcal{L}(x)) \geq 1 - e^{-x} \). Then it holds for any \( m \geq m_0 \) with probability greater \( 1 - 6e^{-x} \)
\[ \| \tilde{D}_m(\theta_m - \theta^*) - \tilde{\xi}_m(\nu^*_m) \| \leq \triangle(r_1, x) + \alpha(m), \]
where \( r_1 > 0 \) is defined in (3.1) and
\[ \tilde{\xi}_m(\nu^*_m) \overset{\text{def}}{=} \tilde{D}_m^{-1}(\nabla_{\theta} - A_m H_m^{-1} \nabla_{\eta}) \mathcal{L}(\nu^*_m). \]

Remark 3.3. Note that we slightly changed the formulation and evaluate the Hessian in \( \nu^* \in l^2 \) instead of in \( (\nu^*_m, 0) \in l^2 \). This means that the matrix \( \tilde{D}_m \in \mathbb{R}^{p^* \times p} \) also is evaluated in \( \nu^* \in l^2 \) instead of in \( (\nu^*_m, 0) \in l^2 \). The proof of the result is the same as the one presented in Andresen and Spokoiny (2014).
It remains to show condition (bias) which reads:

(bias) There exist decreasing functions \(\alpha, \beta : \mathbb{N} \to \mathbb{R}_+\) such that

\[
\|\hat{D}_m(\theta^*_m - \theta^*)\| \leq \alpha(m).
\]

To show that the condition (bias) is met we only have to show that the conditions \((Lr_\infty), (\nu \kappa)\) and \((\kappa)\) from Andresen (2014) are met and then we can use Theorem 2.1 of that work. But exactly this is done in Lemma A.2. In Lemma 3.6 we show that 

\[
\diamondsuit(\tau_0(x), x) \leq \sqrt{x + m}.
\]

This gives for \(m^{-1} \leq (1 - \rho^2) \wedge \frac{1}{2}\) and \(m^{-\alpha+3/2} \leq 1\)

\[
\alpha(m) \leq \sqrt[\rho^2]{1 - \rho^2} \left\{ C_1 2 \sqrt{nm}^{-\alpha-1/2} + 2C\diamondsuit p^{5/2}/\sqrt{n} \right\},
\]

\[
\beta(m) \leq \left\{ \frac{1 + \rho^2}{1 - \rho^2} + 1 \right\} 2m^{-1}
\]

We find

\[
\|\hat{D}_m(\bar{\theta}_m - \theta^*) - \hat{\xi}_m(\nu^*_m)\| \leq \left( 2 + \sqrt[\rho^2]{1 - \rho^2} C\diamondsuit \right) \diamondsuit(r_1, x)
\]

\[
+ \sqrt[\rho^2]{1 - \rho^2} C_1 2 \sqrt{nm}^{-\alpha-1/2}.
\]

For the Wilks theorem we use Theorem 2.6 and Remark 2.12 of Andresen and Spokoiny (2014) where we simply have to plug in our estimates.

**Theorem 3.8.** Assume the same as in Theorem 3.7. Pick a radius \(r_0 > 0\) such that

\[
P^3(\bar{\nu}_m, \bar{\nu}_{\theta^*_m, m} \in Y_{0,m}(r_0)) > 1 - e^{-x},
\]

and set

\[
\delta^*_p \overset{\text{def}}{=} \left( 2 + \frac{1 + \rho}{1 - \rho} \right) \diamondsuit(r_1, x) + \frac{1 + \rho}{1 - \rho} \alpha(m).
\]

Then we get with probability greater \(1 - 6e^{-x}\)

\[
\left| 2\hat{L}(\bar{\theta}_m, \theta^*) - \|\hat{\xi}_m(\nu^*_m)\|^2 \right|
\]

\[
\leq 2 \left\{ r_p + \diamondsuit(r_1 + \delta^*_p, x) \right\} \diamondsuit(r_1 + \delta^*_p, x)
\]

\[
+ \left\{ \delta^*_p + \diamondsuit(r_1 + \delta^*_p, x) \right\}^2 + 2\alpha^2(m) + 2\alpha(m)\delta(x, \hat{B}).
\]
To complete the proof of Proposition 2.2 it remains to ensure that the conditions of
the Corollaries 2.8 and 2.9 of Andresen and Spokoiny (2014) are met. We have to show
that the following two conditions are met:

\begin{enumerate}
  \item[(bias')] As \(m \to \infty\) with \(\| \cdot \|\) denoting the spectral norm
    \[
    \| I - \tilde{D}_m(v^*)^{-1}\tilde{D}(v^*)^2\tilde{D}_m(v^*)^{-1} \| = o(1),
    \]
    \[
    \| I - \tilde{D}_m(v_m^*)^{-1}\tilde{D}_m(v_m^*)^2\tilde{D}_m(v_m^*)^{-1} \| = o(1).
    \]

  \item[(bias'')] Further we need convergence of the covariance of the weighted score. For this define
    \[
    \tilde{\mathbb{V}}^2_{m,2}(v_m^*) \coloneqq \text{Cov} \left( \nabla_{\theta} \ell_1(v_m^*) - A_mH^{-2}\nabla_{\eta_1}(v_m^*) \right).
    \]

\end{enumerate}

For (bias') we use Lemma A.3 of Andresen (2014) combined with Lemma A.2 to
find for \(n \in \mathbb{N}\) large enough

\[
\| I - \tilde{D}_m^{-1}\tilde{D}^2\tilde{D}_m^{-1} \| \leq \sqrt{\frac{1 + \rho^2 + m^{-1}}{1 - \rho^2}} \frac{C_m^2}{C_{D}^2 m^{-1}} \to 0.
\]

As we evaluate the Hessian in \(v^* \in \ell^2\) instead of in \((v_m^*, 0) \in \ell^2\) this already gives
(bias').

Condition (bias'') in our setting becomes

\begin{enumerate}
  \item[(bias'')] The i.i.d. random variables \(Y_i(m) \in \mathbb{R}^p\) satisfy \(\text{Cov}(Y_i(m)) \to 0\) where
    \[
    Y_i(m) \coloneqq \left( \frac{1}{\sqrt{n}} \tilde{D}_m \right)^{-1} \{ \nabla_{\theta} (\ell_i(v_m^*) - \ell_i(v^*)) - A_mH^{-2}\nabla_{(\eta_1, \ldots, \eta_m)} (\ell_i(v_m^*) - \ell_i(v^*)) \}.
    \]

\end{enumerate}

which is proved in the following lemma:

**Lemma 3.9.** Under the conditions of Proposition 2.2 condition (bias'') is satisfied.

Finally we determine an admissible rate for \(m(n) \in \mathbb{N}\) which ensures that the error
terms vanish. By Lemma 3.6 we know that

\[
\Diamond(r_2, x) \lor \Diamond(r_1, x) \leq C(p^* + x)^{5/2}/\sqrt{n}.
\]

If \(p^*5/2/\sqrt{n} \to 0\), we can get that \(2(r_p(x) + \delta_p^*(x))\Diamond(r_2, x) \to 0\) by choosing a sequence
\(x_n > 0\), that increases slow enough. If \(\sqrt{n}m^{-a-1/2} \to 0\) we get the desired result.
Clearly such a sequence exists and in this case \(\mathbb{P}(\Omega(x_n)) \to 1.\)
A Proofs

In the following all the technical steps necessary to prove the Lemmas of section 3 are presented.

A.1 Calculating the elements

First we calculate the relevant objects in this setting. For this we have to emphasize one subtlety about this analysis. As the parameter \( \theta \in \mathbb{R}^p \) lies in \( S^{p,+} \subset \mathbb{R}^p \) a more appropriate parameter set is \( W_S := [0,\pi] \times [-\pi/2,\pi/2] \times [-\pi/2,\pi/2] \times \ldots \times [-\pi/2,\pi/2] \subset \mathbb{R}^{p-1} \). This gives, parametrising the half sphere \( S^{p,+} \subset \mathbb{R}^p \) via the standard spherical coordinates

\[
\Phi : [0,\pi] \times [-\pi/2,\pi/2] \times [-\pi/2,\pi/2] \times \ldots \times [-\pi/2,\pi/2] \subset \mathbb{R}^{p-1} \rightarrow S^{p,+},
\]

that our actual likelihood functional is defined on \( W_S \times \mathbb{R}^m \) as

\[
\mathcal{L}_{m\lambda}(\theta, \eta) = \frac{1}{n} \sum_{i=1}^{n} \left\| Y_i - f_\eta(X_i^\top \Phi(\theta)) \right\|^2 / 2 - n\lambda \| G_1 \eta \|^2 / 2,
\]

where with abuse of notation we denote the preimage of an element of the sphere by the same symbol. Fix any element of the set of maximizers \( \upsilon_m^* \) for some \( m \in \mathbb{N} \). First we calculate

\[
\zeta(\upsilon, \upsilon^*) := \mathcal{L}_{m\lambda}(\upsilon, \upsilon^*) - \mathbb{E}_{\upsilon^*} \mathcal{L}_{m\lambda}(\upsilon, \upsilon^*) = -\sum_{i=1}^{n} \varepsilon_i \left( f_\eta^*(X_i^\top \Phi(\theta^*)) - f_\eta(X_i^\top \Phi(\theta)) \right) + \varepsilon_i^2 / 2.
\]

This gives that with \( \nabla_{\upsilon^*} = (\nabla_{\theta_1}, \ldots, \nabla_{\theta_{p-1}}, \nabla_{\eta_1}, \ldots, \nabla_{\eta_m}) \)

\[
\nabla_{\upsilon^*} \zeta(\upsilon) = \sum_{i=1}^{n} \left( f_\eta'(X_i^\top \Phi(\theta^*)) \nabla \Phi(\theta^*)^\top X_i, e(X_i^\top \Phi(\theta)) \right) \varepsilon_i
\]

\[
def \sum_{i=1}^{n} \varsigma_{i,m}(\upsilon) \varepsilon_i \]

\[
def W_m(\upsilon) \varepsilon.
\]

where with \( e = (e_1, \ldots, e_m) \)

\[
W_m(\upsilon) = \begin{pmatrix}
  f_\eta'(X_1^\top \theta) \nabla \Phi(\theta)^\top X_1 & \ldots & f_\eta'(X_n^\top \theta) \nabla \Phi(\theta)^\top X_n \\
  e(X_1^\top \theta) & \ldots & e(X_n^\top \theta)
\end{pmatrix}.
\]

By assumption the \( \varepsilon_i \) are i.i.d. with covariance \( \sigma^2 > 0 \) and the design points \( (X_i) \)
are i.i.d. as well. We get

\[ \gamma_m^2 \overset{\text{def}}{=} \sigma^2 \mathbb{E} W_m(v^*) W_m(v^*)^\top = n\sigma^2 \begin{pmatrix} d_0^2(v^*) & a_m(v^*) \\ a_m(v^*) & h_m^2(v^*) \end{pmatrix} \overset{\text{def}}{=} n\sigma^2 d_m^2 \in \mathbb{R}^{(p-1+m) \times (p-1+m)}. \]

where with \( \mathbb{E}[\cdot] \) the expectation under the measure \( \mathbb{P}^X_1 \)

\[ d_0^2(v) = \mathbb{E}\left[ f'_{\eta}(X_1^\top \theta)^2 \nabla \Phi(\theta)^\top X_1 X_1^\top \nabla \Phi(\theta) \right], \]
\[ a_m^2(v) = \mathbb{E}\left[ e e^\top (X_1^\top \theta) \right], \]
\[ h_m(v) = \mathbb{E}\left[ f'_{\eta}(X_1^\top \theta) \nabla \Phi(\theta)^\top X_1 e^\top (X_1^\top \theta^*) \right]. \]

Further we get because of the quadratic functional and sufficient smoothness of the basis \( (e_i) \) we have for any \( v \in \mathbb{R}^{p-1} \)

\[ -\nabla_p^2 \mathbb{E}[\mathcal{L}_{m \lambda}(v)] \overset{\text{def}}{=} \mathcal{D}_m^2(v) = n a_m^2(v) + nr_m^2(v), \]
\[ n a_m^2 = n \begin{pmatrix} d_0^2(v) & a_m(v) \\ a_m(v) & h_m^2(v) \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} D(v)^2 & A_m(v) \\ A_m(v) & H_m^2(v) \end{pmatrix}, \]
\[ r_m^2(v) = \mathbb{E}\left[ (f_{\eta}(X^\top \theta) - f_{\eta^*}(X^\top \theta^*)) \begin{pmatrix} v_0^2(v) \\ b_m(v) \end{pmatrix} \right], \]
\[ v_0^2(v) = 2f''_{\eta}(X^\top \theta) \nabla \Phi^T_0 X(X)^\top \nabla \Phi_0 + |f'_{\eta}(X^\top \theta)|^2 \nabla^2 \Phi^T_0 [X, \cdot, \cdot], \]
\[ b_m(v) = \nabla \Phi_0 X^\top e^\top (X^\top \theta), \]

such that

\[ \mathcal{D}_{m \lambda}^2(v^*) = n d_m^2(v) \in \mathbb{R}^{(p-1+m) \times (p-1+m)}. \]

### A.2 Proof or Lemma 3.2

First we proof the following 2 lemmas

**Lemma A.1.** We have \((\mathcal{L}_r)\) with \( b(r) \equiv c(r) \)

**Proof.** The proof for \((\mathcal{L}_r)\) nearly matches the one for \((\mathcal{L}_x)\) in Lemma A.12 the only difference being that the term \( -n \mathbb{E}[\|f_{\eta_m}(X^\top \theta^*_m) - f_{\theta}(X^\top \theta^*)\|^2 |X] \) does not show up. \( \square \)
Lemma A.2. Assume that the density \( f_X \) is Lipshitz continuous and that the \( X \in \mathbb{R} \) are bounded by some constant \( s_X > 0 \). Then using our orthogonal and sufficiently smooth wavelet basis we get for any \( \lambda \in [0, 1] \)

\[
\|\mathcal{H}_m^{1/2} x^\ast\|^2 < \left( C(m) \| f_X^\top \theta^\ast \|_\infty C_{\| f^\ast \|} + C(m) 26\sqrt{28} s_X^{p+1} L_{f_X} \| \psi^\ast \|_\infty C_{\| f^\ast \|} \right) n m^{-2\alpha},
\]

\[
\alpha(m) \overset{\text{def}}{=} \| \mathcal{D}_m^{-1} \mathcal{A}_{v^\ast} x^\ast \| \leq C_1 m^{-\alpha - 1/2} \sqrt{n},
\]

\[
\beta(m) \overset{\text{def}}{=} \| \mathcal{D}_m^{-1} \mathcal{A}_{v^\ast} H_m \| \leq \frac{2C_1}{c_D} m^{-1/2},
\]

\[
\tau(m) \overset{\text{def}}{=} \| \mathcal{D}_m^{-1} \nabla_{v^\ast} \mathbb{E}[\mathcal{L}(v^\ast, \lambda x^\ast) - \mathcal{A}_{v^\ast}] x^\ast \| \leq C_1 m^{-2\alpha + 1/2} \sqrt{n},
\]

\[
0 = \left[ x^\ast \top (\mathcal{H}_m - \nabla_{v^\ast} \mathbb{E}[\mathcal{L}(v^\ast, \lambda x^\ast)]) x^\ast \right],
\]

where

\[
C_1 \leq 26\sqrt{28 \sqrt{(2\alpha - 3)/(2\alpha - 4)}} C(m)^{3/2}
\]

\[
\left( \frac{\sqrt{p + 2}}{2} \pi \| \psi^\prime \|_\infty s_X^2 \| d_X \theta \|_\infty C_{\| f^\ast \|} + C_{\| f^\ast \|} s_X^{p+1} L_{f_X} \| \psi^\ast \|_\infty \right)
\]

\[
\sqrt{13} \sqrt{28C(m)^2 s_X^{p+1} L_{f_X} \| \psi^\ast \|_\infty C_{\| f^\ast \|}^2}
\]

\[
C(m) = \frac{j_m 12 + 2j_m - 1}{2j_m} \approx 1.
\]

Further we find that

\[
\| D \| \leq \frac{p + 2}{4} C_{\| f^\ast \|} \| \psi^\prime \|_\infty s_X^2 \pi^2.
\]

Proof. We have that

\[
\| \mathcal{D}_m^{-1} \mathcal{A}_{v^\ast} x^\ast \| \leq \| \mathcal{D}_m^{-1} \| \| \mathcal{A}_{v^\ast} x^\ast \|.
\]

We will see in Lemma A.4 that

\[
\| \mathcal{D}_m^{-1} \| \leq \frac{1}{c_D \sqrt{n}}.
\]

And we have by definition that for any \( v = (\theta, \eta) \in \mathbb{R}_S \times \mathbb{R}^m \)

\[
\frac{1}{n} | v^\top \mathcal{A}_{v^\ast} x^\ast | \leq \frac{1}{n} | \theta A_{\theta^\ast} x^\ast | + \frac{1}{n} | \eta A_{\eta^\ast} x^\ast |.
\]

We first analyze the second summand

\[
\frac{1}{n} \eta A_{\eta^\ast} x^\ast = \sum_{l=m+1}^{\infty} \eta^\prime \sum_{k=1}^m \eta_k \mathbb{E}[e_k e_l (X^\top \theta^\ast)].
\]
Observe that if the density of \( f_X : \mathbb{R}^p \mapsto \mathbb{R} \) is Lipshitz continuous with Lipshitz constant \( L_{f_X} \) and its support contained in a ball of Radius \( s_X > 0 \) then the density \( f_{X^\top \theta^*} : \mathbb{R} \mapsto \mathbb{R} \) of \( X^\top \theta^* \in \mathbb{R} \) is Lipshitz continuous with Lipshitz constant \( L_{f_{X^\top \theta^*}} \leq s_X^p L_{f_X} \).

Further for \( k, l \in \mathbb{N} \)

\[
\mathbb{E}[e_k e_l (X^\top \theta^*)] = \int_{[-s_X, s_X]} e_k(x) e_l(x) f_{X^\top \theta^*}(x) dx.
\]

Denote by \( I_k \subset \mathbb{R} \) the support of \( e_k(x) \). We write

\[
\mathbb{E}[e_k e_l (X^\top \theta^*)] = \int_{I_k} e_k(x) e_l(x) f_{X^\top \theta^*}(x) dx
\]

\[
= \int_{I_k} e_k(x) e_l(x) f_{X^\top \theta^*}(x_0) dx 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l)
\]

\[
+ \int_{I_k} e_k(x) e_l(x) \left( f_{X^\top \theta^*}(x) - f_{X^\top \theta^*}(x_0) \right) dx 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l),
\]

where \( x_0 \in I_l \) is the center of the support of \( e_l(x) \), which is of length \( 2^{-j_l} 26s_X \) for \( l = j_l 12 + 2^j + r_l \in \mathbb{N} \). Because of orthogonality the first summand on the right hand side is equal to zero. For the second summand we use the Lipshitz continuity and Cauchy Schwarz to estimate

\[
| \int_{I_k} e_k(x) e_l(x) \left( f_{X^\top \theta^*}(x) - f_{X^\top \theta^*}(x_0) \right) dx 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l) |
\]

\[
\leq s_X^p L_{f_X} 2^{-j_l - 1} \int_{I_k} |e_k(x)| |e_l(x)| dx 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l)
\]

\[
\leq s_X^p L_{f_X} 2^{-j_l - 1} \left( \int_{I_k} |e_l(x)|^2 dx \right)^{1/2} 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l)
\]

\[
\leq s_X^p L_{f_X} 2^{-j_l - 1} \left( \int_{I_k} |e_k(x)|^2 dx \right)^{1/2} 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l)
\]

\[
\leq 26 s_X^{j_l + 1} L_{f_X} \| \psi \|_\infty 2^{-j_l - 1} 2^{j_l/2 - j_l/2} 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l),
\]

where we used that the \( (e_k) \) form an orthonormal basis, that \( \| e_k \|_\infty \leq 2^{j_k/2} \| \psi \|_\infty \) and that \( I_l \) is of length \( 2^{-j_l} 26s_X \). Note that for each \( j_k = 0, \ldots, j_m \) there exists at most 13 \( r_k(l) \in \{0, \ldots, 2^{j_k} + 1\} \) with \( I_l \cap I_k \neq \emptyset \). Remember that \( m = j_m 12 + 2^{j_m} - 1 \) and
note that $2^{jm} \leq m$. This gives using the Cauchy Schwartz inequality and that $\|\eta\| = 1$

\[
\frac{1}{n} \eta A_{\eta \times x^*} \leq 26 s_x^{p+1} L_{fx}\|\psi^2\|_\infty \sum_{l=m+1}^{\infty} \sum_{k=1}^{m} |\eta_k^*|^2 |\eta_k||2^{-j_l-1} 2^{j_k/2} s_{l,(I_l \cap I_k \neq \emptyset)}(k,l) \]

\[
\leq 26 s_x^{p+1} L_{fx}\|\psi^2\|_\infty \sum_{l=m+1}^{\infty} |\eta_k^*|^2 |2^{-3j_l/2} \left( \sum_{k=1}^{m} 2^{j_k} 1_{(I_l \cap I_k \neq \emptyset)}(k,l) \right)^{1/2} \]

\[
\leq 26 \sqrt{13} s_x^{p+1} L_{fx}\|\psi^2\|_\infty \sqrt{m} \left( \sum_{l=m+1}^{\infty} 2^{-3j_l} \right)^{1/2} \left( \sum_{l=m}^{\infty} |\eta_l^*|^2 \right)^{1/2} \left( \sum_{l=m}^{\infty} 2^{-3j_l} \right)^{1/2} . \]

By assumption \textbf{Cond}_\nu,*

\[
\left( \sum_{l=m+1}^{\infty} |\eta_l^*|^2 \right)^{1/2} \leq m^{-\alpha} \left( \sum_{l=m+1}^{\infty} l^{2\alpha} |\eta_l^*|^2 \right)^{1/2} \leq m^{-\alpha} c_{\|\eta\|} .
\]

Since $m = j_m 12 + 2^{j_m} - 1$ and $l = j_l 12 + 2^l + r_l$ with $r_l \in \{0, \ldots, 2^{j_l} + 1\}$

\[
\left( \sum_{l=m+1}^{\infty} 2^{-3j_l} \right)^{1/2} = \left( \sum_{j_l=j_m}^{\infty} C(m) 2^{j_l} 2^{-3j_l} \right)^{1/2} = C(m)^{1/2} 2^{-j_m} 2 \leq \sqrt{2} C(m)^{3/2} m^{-1} ,
\]

with

\[
C(m) = \frac{j_m 12 + 2^{j_m} - 1}{2^{j_m}}.
\]

This gives

\[
\frac{1}{n} \eta A_{\eta \times x^*} \leq \sqrt{2} C(m)^{3/2} c_{\|\eta\|}^{2} 26 \sqrt{13} s_x^{p+1} L_{fx}\|\psi^2\|_\infty m^{-\alpha-1/2} .
\]

Similarly we get by the dominated convergence theorem

\[
\frac{1}{n} \theta A_{\theta \times x^*} = \sum_{k=1}^{\infty} \sum_{l=m+1}^{\infty} \eta_k^* \eta_l^* E[(X^\top \nabla \phi(\theta^*) \theta) e_l' e_l (X^\top \theta^*)] 1_{(I_l \cap I_k \neq \emptyset)}(k,l).
\]
To justify the exchange of summation and expectation note that for each $l \in \mathbb{N}$

\[
\mathbb{E} \left[ \left| (X^\top \nabla \Phi(\theta^*) e_l f'_{\eta^*}(X^\top \theta^*)) \right| \right] \\
\leq \left\| \nabla \Phi(\theta^*) \right\|_{sX} 2^{\frac{l}{2}} \mathbb{E} \left[ \left| f'_{\eta^*}(X^\top \theta^*) \right| \right] \\
\leq \left\| \nabla \Phi(\theta^*) \right\|_{sX} 2^{\frac{l}{2}} \mathbb{E} \left[ \left| \sum_{k=1}^{\infty} \eta_k e_k'(X^\top \theta^*) \right| \right] \\
\leq \left\| \nabla \Phi(\theta^*) \right\|_{sX} 2^{\frac{l}{2}} \left( \sum_{k=1}^{\infty} \eta_k e_k'(X^\top \theta^*) \right)^{1/2} \left( \sum_{k=1}^{\infty} l^{-2\alpha} 2^{3j_k} \left\| \psi' \right\|_2^2 \right)^{1/2} \\
\leq \left\| \nabla \Phi(\theta^*) \right\|_{sX} 2^{\frac{l}{2}} \left( \frac{13}{2} \sum_{j=0}^{\infty} l^{-2\alpha} 2^{4j} \right)^{1/2} < \infty.
\]

The exchange of the order of summation is justified by the subsequent bounds and again the dominated convergence theorem. Using that for any $\theta \in W_S$ it holds true that $\left\| \nabla \Phi(\theta^*) \theta \right\| \leq \frac{\sqrt{p+2}}{2} \pi$ we estimate similarly to before

\[
\mathbb{E} \left[ \left| (X^\top \theta) e_k e_l (X^\top \theta^*) \right| \right] \\
\leq \frac{\sqrt{p+2}}{2} \pi s_{\mathbb{X}}^2 \mathbb{E} \left[ \left| e_k e_l (X^\top \theta^*) \right| \right] \\
\leq \frac{\sqrt{p+2}}{2} \pi s_{\mathbb{X}}^2 \int_{l_i} e_k'(x) e_l(x) dX^\top \theta^* (x) dx \\
\leq \frac{\sqrt{p+2}}{2} \pi s_{\mathbb{X}}^2 \left\| dX^\top \theta \right\|_{\infty} \left( \int_{l_i} e_k'(x)^2 dx \right)^{1/2} \left( \int_{l_i} e_l(x)^2 dx \right)^{1/2} \\
\leq 26 \frac{\sqrt{p+2}}{2} \pi \left\| \psi' \right\|_{\infty} s_{\mathbb{X}}^2 \left\| dX^\top \theta \right\|_{\infty} 2^{3j_k/2} \left( j_i \vee j_k \right)^{1/2} 2^{-j_i} \left( 1 \{ l_i \cap l_k \neq 0 \} \right) (k, l).
\]

Together this gives with similar arguments as above

\[
\frac{1}{n} | \theta A_{\theta^*, \kappa^*} | \\
\leq 26 \frac{\sqrt{p+2}}{2} \pi \left\| \psi' \right\|_{\infty} s_{\mathbb{X}}^2 \left\| f_X \right\|_{\infty} \sum_{k=1}^{\infty} \eta_k k^{3/2} \sum_{l=m+1}^{\infty} \eta_l^{3/2} \left( j_i \vee j_k \right)^{1/2} 2^{-j_i} \left( 1 \{ l_i \cap l_k \neq 0 \} \right) (k, l) \\
\leq 26 \frac{\sqrt{p+2}}{2} \pi \left\| \psi' \right\|_{\infty} s_{\mathbb{X}}^2 \left\| f_X \right\|_{\infty} \sum_{k=1}^{\infty} \eta_k k^{3/2} \left( \sum_{l=m+1}^{\infty} l^{2\alpha} \eta_l^{2} \right)^{1/2} \\
\left( \sum_{j_i = j_{m+1}}^{\infty} \sum_{j_k = j_{m+1}}^{\infty} 2^{-2\alpha j_i} 2^{-2\alpha j_k} \left( 1 \{ l_i \cap l_k \neq 0 \} \right) (k, l) \right)^{1/2}.
\]
Now we have due to (2.4) that

\[
\sum_{j_l=j_m+1}^{\infty} \sum_{r_l=0}^{2^l+1} 2^{-2\alpha j_l} 2^{-(j_l \vee j_l)} 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l) \\
= \sum_{j_l=j_m+1}^{\infty} 2^{-2\alpha j_l} 2^{-(j_l \vee j_l)} \sum_{r_l=0}^{2^l+1} 1_{\{I_l \cap I_k \neq \emptyset\}}(k, l) \\
= \sum_{j_l=j_m+1}^{\infty} 2^{-2\alpha j_l} 2^{-(j_l \vee j_l)} \left\{ l = j_l 12 + 2^l + r_l \mid r_l \in \{0, \ldots, 2^l + 11\}, I_l \cap I_k \neq \emptyset \right\} \\
= \sum_{j_l=j_m+1}^{\infty} 2^{-2\alpha j_l} 2^{-(j_l \vee j_l)} + 2^{(j_l-j_l) \vee (j_l-j_l) \vee (j_l-j_l)} 2^{(j_l-j_l) \vee (j_l-j_l) \vee (j_l-j_l)} \leq 2^{-\alpha m} 14 \leq C(m) m^{-(\alpha+1)/2}.
\]

Which gives

\[
\frac{1}{n} |\theta A_{\theta, x} x^*| \\
\leq 26 \sqrt{14} C(m)^{1/2} \frac{\sqrt{p^2 + 2}}{2} \pi \|\psi_f\|_{\infty} s_{f_X}^2 \|f_X\|_{\infty} C_{f^*} \sum_{k=1}^{\infty} \eta_k^{2k^2 \alpha} \left( \sum_{k=1}^{\infty} k^{-2(\alpha-3)} \right)^{1/2} \\
\leq 26 \sqrt{14} C(m)^{1/2} \frac{\sqrt{p^2 + 2}}{2} \pi \|\psi_f\|_{\infty} s_{f_X}^2 \|f_X\|_{\infty} C_{f^*} \left( \frac{\sqrt{(2\alpha - 3)/(2\alpha - 4)}}{m^{-(\alpha+1)/2}} \right),
\]

since \( \alpha > 2 \) such that \( \sum_{k=1}^{\infty} k^{-2(\alpha-3)} < (2\alpha - 3)/(2\alpha - 4) \). Collecting both summands we find

\[
\|D_m^{-1} A_{\theta, x} x^*\| \leq c_1 m^{-(\alpha+1)/2} \sqrt{n},
\]

with \( c_1 > 0 \) from (A.1). The same arguments give

\[
\|D_m^{-1} A_{\theta, x} H_m^{-1}\| \leq \frac{1}{c_2} \left( \sup_{\|\theta\|=1, \|x\|=1} \frac{1}{n} |\theta A_{\theta, x} x^*| + \sup_{\|\eta\|=1, \|x\|=1} \frac{1}{n} \|\eta A_{\eta, x} x^*\| \right) \\
\leq \frac{c_1}{c_2} 2 m^{-1/2}.
\]

We bound using the dominated convergence theorem (applicable due to similar bounds as above)

\[
\|H_m x^*\|^2 \leq n \sum_{k=m+1}^{\infty} \eta_k^2 \|f_X^\top \theta^*\|_{\infty} + 2n \sum_{l>k} \eta_k \eta_l^\top E[e_k e_l (X^\top \theta^*)]. \tag{A.2}
\]
As above we find

\[ |IE[e_k e_l (X^\top \theta^*)]| \leq 26 s_{X_{ij}}^{p+1} L_{fx} \|\psi\|_\infty 2^{-3j_l} 2^{-1} 2^{j_k/2} 1_{\{I_l \cap I_k \neq \emptyset\}} (k, l). \]

We estimate

\[
\sum_{l>k>m} \eta_l^* \eta_k^* IE[e_k e_l (X^\top \theta^*)] \\
\leq 26 s_{X_{ij}}^{p+1} L_{fx} \|\psi\|_\infty \sum_{l>k} \eta_l^* \eta_k^* 2^{-3j_l} 2^{-1} 2^{j_k/2} 1_{\{I_l \cap I_k \neq \emptyset\}} (k, l) \\
\leq 26 s_{X_{ij}}^{p+1} L_{fx} \|\psi\|_\infty \sum_{k=1}^\infty \eta_k^* 2^{j_k/2} \sum_{l=k+1}^\infty \eta_l^* 2^{-3j_l} 2^{-1} 1_{\{I_l \cap I_k \neq \emptyset\}} (k, l) \\
\leq 26 s_{X_{ij}}^{p+1} L_{fx} \|\psi\|_\infty \sum_{k=1}^\infty \eta_k^* 2^{j_k/2} \left( \sum_{l=k+1}^\infty \eta_l^* 2^{2\alpha} \right)^{1/2} \\
\left( \sum_{l=k+1}^\infty l^{-2\alpha} 2^{-3j_l} 1_{\{I_l \cap I_k \neq \emptyset\}} (k, l) \right)^{1/2}.
\]

We continue using that \( l \geq 2^j \)

\[
\sum_{l=k+1}^\infty l^{-2\alpha} 2^{-3j_l} 1_{\{I_l \cap I_k \neq \emptyset\}} (k, l) \\
\leq \sum_{l=k+1}^\infty 2^{-(3+2\alpha)j_l} 1_{\{I_l \cap I_k \neq \emptyset\}} (k, l) \\
\leq \sum_{j=jk+1}^\infty 2^{-(3+2\alpha)j} \{|l = 12j + 2j, \ldots, 12j + 2^{j+1} + 11 : I_l \cap I_k \neq \emptyset\}| \\
= \sum_{j=jk+1}^\infty 2^{-(3+2\alpha)j} 2^{2j-j_k} 13^{-} \\
\leq 2^{-j_k} 14 \sum_{j=jk+1}^\infty 2^{-(2+2\alpha)j} \\
= 2^{-(3+2\alpha)j_k} 14 \sum_{j=0}^\infty 2^{-(2+2\alpha)j} \leq 2^{-(3+2\alpha)j_k} 28.
\]
Plugging this in we find

\[
\sum_{l > k > m} \eta_l^* \eta_k^* \mathbb{E}[e_k e_l (X^\top \theta^*)]
\]

\[
\leq 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty \sum_{k=m+1}^\infty \eta_k^* 2^{-(2+2\alpha)j_k} / 2 C_{f^*} \|f^*\|^{1/2}
\]

\[
\leq 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*} \|f^*\| \left( \sum_{k=m+1}^\infty \eta_k^* 2^{2\alpha k} \right)^{1/2} \left( \sum_{k=m+1}^\infty k^{-2\alpha} 2^{-(2+2\alpha)j_k} \right)^{1/2}
\]

\[
\leq 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*}^2 \|f^*\|^2 \left( \sum_{j=m}^\infty 2^{-(1+4\alpha)j} \right)^{1/2} \left( \sum_{j=0}^\infty 2^{-(1+4\alpha)j} \right)^{1/2}
\]

From which we obtain

\[
\|H_m \eta_2^*\|^2 = n \sum_{k=m+1}^\infty \eta_k^* 2^{2\alpha k} \|f_{X^\top \theta^*}\|_\infty + 2 s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*}^2 \|f^*\|^{1/2} n 2^{-(1+4\alpha)(j_m+1)/2}
\]

\[
\leq \|f_{X^\top \theta^*}\|_\infty n m^{-1/2} 2^{2\alpha k} \left( \sum_{k=m+1}^\infty \eta_k^* 2^{2\alpha k} \right) + C(m) 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*}^2 \|f^*\|^{1/2} \leq C(m) 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*}^2 \|f^*\|^{1/2}
\]

\[
\leq \left( C(m) \|f_{X^\top \theta^*}\|_\infty C_{f^*} \right) n m^{-1/2} \left( \sum_{k=m+1}^\infty \eta_k^* 2^{2\alpha k} \right) + C(m) 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*}^2 \|f^*\|^{1/2} \leq C(m) 26 \sqrt{28} s_X^{p+1} L_{f_X} \|\psi\|_\infty C_{f^*}^2 \|f^*\|^{1/2} n m^{-1/2}
\]

Next we show

\[
\|D_m^{-1} (\nabla_{\psi, \lambda} \mathbb{E}[\mathcal{L}(\psi^*, \lambda \lambda^*)] - A_{\psi, \lambda}) \lambda^*\| \leq \tau(m).
\]

For this note that

\[
(\nabla_{\psi, \lambda} \mathbb{E}[\mathcal{L}(\psi^*, \lambda \lambda^*)] - A_{\psi, \lambda}) \lambda^*
\]

\[
= n \begin{pmatrix} \mathbb{E}[f_{(0, \lambda \lambda^*)} f_{(0, \lambda \lambda^*)} (X^\top \theta^*)] \left( \mathbb{E}[f_{(0, \lambda \lambda^*)} f_{(0, \lambda \lambda^*)} (X^\top \theta^*)] \right) + n \left( \mathbb{E}[f_{(0, \lambda \lambda^*)} f_{(0, \lambda \lambda^*)} (X^\top \theta^*)] \right) \end{pmatrix}.
\]
We infer

\[
\|D_m^{-1}(\nabla \psi \mathbb{E}[\mathcal{L}((\psi^*, \lambda^* \theta^*))]) - A_{\psi^*} \| \leq \left( \mathbb{E} \left[ \left\| D_m^{-1} \left( \frac{f'(0, \lambda^*)}{e} \right) \right\|^2 \right] \right)^{1/2} + \mathbb{E} \left[ \left\| D_m^{-1} \left( \frac{f'(0, \lambda^*)}{0} \right) \right\|^2 \right]^{1/2}
\]

\[
\leq \frac{\sqrt{n}}{eD} \left\{ \mathbb{E}[f'(0, \lambda^*)]^2 \right\}^{1/2} + \mathbb{E}[f'(0, \lambda^*)]^2 + \|d_{X^* \theta}\|^{1/2} C(m)^{1/4} \sqrt{m}
\]

\[
\mathbb{E}[f'(0, \lambda^*)^2] \leq \lambda \sum_{k,l,m+1} \eta_k^* \eta_l^* \mathbb{E}[e_k^* e_l^*(X^* \theta^*)]
\]

\[
\leq 26s \|\psi\|_{\infty} \|f\|_{\infty} \sum_{k,l,m+1} \eta_k^* \eta_l^* 2^{3j_l + j_k}/2^{-2(j_l \vee j_k)} 1_{I_k \cap I_l \neq \emptyset}(k, l)
\]

\[
\leq 26s \|\psi\|_{\infty} \|f\|_{\infty} \sum_{k,m+1} \eta_k^* 2^{3j_k}/2 \sum_{l=m+1} \eta_l^* 2^{3j_l}/2^{-2(j_l \vee j_k)} 1_{I_k \cap I_l \neq \emptyset}(k, l)
\]

\[
\leq 26s \|\psi\|_{\infty} \|f\|_{\infty} \sum_{k,m+1} \eta_k^* 2^{3j_k}/2 \left( \sum_{l=m+1} 2^{2\alpha j_l^*} \right)^{1/2}
\]

\[
\left( \sum_{l=m+1} 2^{(3-2\alpha) j_l - 2(j_l \vee j_k)} 1_{I_k \cap I_l \neq \emptyset}(k, l) \right)^{1/2}
\]

Observe

\[
\sum_{l=m+1} 2^{(3-2\alpha) j_l - 2(j_l \vee j_k)} 1_{I_k \cap I_l \neq \emptyset}(k, l)
\]

\[
= \sum_{j=j_m+1} 2^{(3-2\alpha) j - 2(j \vee j_k)} \left| \left\{ l = j12 + 2^j + r_l \mid r_l \in \{0, \ldots, 2^j + 11\}, I_l \cap I_k \neq \emptyset \right\} \right|
\]

\[
= \sum_{j=j_m+1} 2^{(3-2\alpha) j - 2(j \vee j_k) \leq 14} \sum_{j=j_m+1} 2^{2(2-2\alpha) j} = C(m)^3 14m^{-2\alpha + 2}.
\]
Such that again using the Cauchy Schwarz inequality for any $\lambda \in [0, 1]$

$$IE[f'_{(0, \lambda \varphi^*)}^2] \leq 26C(m)^{3/2} \sqrt{14sX} \|\psi'\|_\infty \|f_X\|_\infty \sigma^{\alpha+1} \sum_{k=m+1}^\infty \eta_k \sigma^{23k/2}$$

$$\leq 26C(m)^2 \sqrt{14sX} \|\psi'\|_\infty \|f_X\|_\infty \sigma^2 \sigma^{m-2\alpha+3}.$$ 

Further

$$IE[f^2_{(0, \lambda \varphi^*)}(X^T \theta^*)] = \sum_{k,l=m+1}^\infty \eta_k \eta_l IE[e_k e_l(X^T \theta^*)]$$

$$\leq 13s^{p+1} X f_X \|\psi\|_\infty \sum_{k,l=m+1}^\infty \eta_k \eta_l \sigma^{2^{3(ji\vee jk)/2 + (ji \wedge jk)/2}} 1_{(k \cap l \neq \emptyset)}(k, l)$$

$$= 13s^{p+1} X f_X \|\psi\|_\infty \sum_{k=m+1}^\infty \eta_k \sigma^{2^{j-ik}} \sum_{l=m+1}^\infty \eta_l \sigma^{2^{j-ik}} 1_{(k \cap l \neq \emptyset)}(k, l)$$

$$\leq 13s^{p+1} X f_X \|\psi\|_\infty \sum_{k=m+1}^\infty \eta_k \sigma^{2^{j-ik}} \left( \sum_{l=m+1}^\infty \sigma^{2\alpha j 14} \right)^{1/2}$$

$$\leq 13 \sqrt{14C(m)^{1/2}} L_X f_X \|\psi\|_\infty \sigma^{2^{\alpha-ik}} \sum_{k=m+1}^\infty \eta_k \sigma^{2^{j-ik}}$$

$$\leq 13 \sqrt{28C(m)^{2}} L_X f_X \|\psi\|_\infty \sigma^{m-2\alpha} \sum_{k=m+1}^\infty \eta_k \sigma^{2^{j-ik}}$$

Together this gives

$$\|D_m^{-1} (\nabla_{\varphi^*} IE[L(\varphi^*, \lambda \varphi^*)]) - A_{\varphi^*} \| \varphi^* \|$$

$$\leq \frac{1}{cD} \left( 2sX \left\{ 26C(m)^{2} \sqrt{14sX} \|\psi'\|_\infty \|f_X\|_\infty \sigma^2 \|f^*\|_\infty \right\}^{1/2} + \|d_X \|_\theta^{1/2} C(m)^{1/4} \right)$$

$$\sqrt{\frac{13 \sqrt{28C(m)^{2}} L_X f_X \|\psi\|_\infty \sigma^{m-2\alpha+1/2} \sqrt{n}}{n}}$$

$$\leq c_1 m^{-2\alpha+1/2} \sqrt{n}.$$ 

Clearly

$$\varphi^* \top (\mathcal{H}_m - \nabla_{\varphi^*} IE[L(\varphi^*, \lambda \varphi^*)]) \varphi^* = 0.$$
To see this simply note that for any $f \in S$ and any $\varpi \in S$

$$\varpi^T \nabla_{\varpi} \mathbb{E} L(\theta^*, f \varpi) = \mathbb{E} [f^2_{(0, \varpi)}(X^T \theta^*)] = \varpi^T \mathcal{H}_m \varpi.$$  

Further we find that

$$\theta^T D \theta = \mathbb{E} [f^2_f(X^T \theta^*)] \leq \|f\|_\infty^2 s_X^2 \leq \frac{p + 2}{4} c_{\|f\|_\infty} \|\psi\|_\infty^2 s_X^2 \pi^2,$$

and the proof is complete. \hfill \Box

Now the claim of Lemma 3.2 is proved by an application of the next Lemma, which is Lemma A.1 of Andresen (2014).

**Lemma A.3.** Assume that $(\mathcal{L}_{T, \infty})$ is satisfied with $b(r) \equiv b$ and that the condition $(\varpi)$ is satisfied. Then we get $\|D(v^*_m - v^*)\| \leq r^*$ where $r^* = 4C_{\varpi} \cdot m/b$.

### A.3 Proof of Lemma 3.3

Before we prove the claims of the lemma we need a series of auxiliary lemmas.

**Lemma A.4.** (Cond$_{X, \rho}$) gives $D_m \geq \sqrt{n} c_D$ with

$$c_D \geq \lambda_{\min}(\mathcal{H}) c_{f_\eta}^2 c_{\mathcal{F} \eta}^2 \pi^2 \|\theta^*\|^2 / \left(8s_X C_{f_\eta} \|\eta\|_\infty\right) \wedge c_{f_\eta}^2 c_{\mathcal{F} \eta}^2 \pi^2 / 2 \wedge \lambda_{\min}(\mathcal{H}) / 4,$$

independent of $m, n$.

**Remark A.1.** We assume that the density of the regressors satisfies $f_X \geq c_{f_X} > 0$. This gives that the density of $X^T \theta^*$ is also bounded away from zero by a constant $c_{f_X \theta^*} > 0$. As we use a orthonormal wavelet basis on $L^2([-s_X, s_X])$ this gives

$$\lambda_{\min}(\mathcal{H}) \geq \inf_{\eta \in l^2} \mathbb{E} \left[f_\eta(X^T \theta^*)^2 \right] / \|\eta\|^2 \geq c_{f_X \theta^*} \int_{[-s_X, s_X]} f_\eta(x)^2 dx / \|\eta\|^2 = c_{f_X \theta^*}.$$  

**Proof.** Observe that for $\nu = (\theta, \eta) = (\nabla \Phi(\theta^*) \gamma_\theta, \eta) \in \mathbb{R}^{p+1}$ where $\gamma_\theta \in [0, \pi] \times [-\pi/2, \pi/2]^{p-2} \subset \mathbb{R}^{p-1}$ and $\theta = \nabla \Phi(\theta^*) \gamma_\theta \in (\theta^*)^\perp$. First assume that $\|\theta\| \leq \lambda_{\min}(\mathcal{H}) / (8s_X C_{f_\eta} \|\eta\|_\infty) \wedge 1/2$ then $\|\eta\| > 1/2$ such that

$$\nu^T D_m \nu = n \mathbb{E} \left[f_\eta(X^T \theta^*)^2 \right] \geq n \mathbb{E} \left[f_\eta(X^T \theta^*)^2 \right] \geq n \lambda_{\min}(\mathcal{H}) \|\eta\| - 2ns_X C_{f_\eta} \|\eta\| \geq \lambda_{\min}(\mathcal{H}) / 2.$$
Otherwise for $\|\theta\| \geq \lambda_{\min}(H)/(4sX C_{f^*,\infty}) ^{1/2}$
\[
v^T D_m v = nE\left(f^*_\eta(x)X^T \theta + f^*_{\eta}(X^T \theta)\right)^2
\]
\[
= nE \left[ E \left( f^*_\eta(x)X^T \theta + f^*_{\eta}(X^T \theta) \bigg| X^T \theta = x \right) \right]
\]
\[
\geq nE \left[ \text{Var} \left( f^*_\eta(x)X^T \theta + f^*_{\eta}(X^T \theta) \bigg| X^T \theta = x \right) \right]
\]
\[
= n\|\theta\|E \left[ f^*_\eta(X^T \theta)^2 \text{Var} \left( X^T \theta / \|\theta\| \bigg| X^T \theta \right) \right].
\]
Now by assumption (Cond$_{X\theta^*}$) it holds true that $\|f^*_\eta(X^T \theta^*)\| > c_{f^*_\eta} > c_{f\theta^*}$ for some $c_{f^*_\eta}, c_{f\theta^*} > 0$. Further by (Cond$_X$) we have that $\text{Var} \left( X^T \theta / \|\theta\| \bigg| X^T \theta \right) > \sigma_{X\theta^*}^2$. This gives
\[
v^T D_m v \geq n\lambda_{\min}(H)c_{f^*_\eta}^2 c_{f\theta^*}^2 \sigma_{X\theta^*}^4 / (8sX C_{f^*,\infty}^2) ^{1/2} \geq n\lambda_{\min}(H)c_{f^*_\eta}^2 c_{f\theta^*}^2 \sigma_{X\theta^*}^4 / 2.
\]
The same argument works for the full operator $D^2$.

Remember that
\[
s_i,m(v^*_m) \overset{\text{def}}{=} \left( f^*_\eta(X_i^T \theta^*_m)\nabla \phi(\theta^*_m)^T X_i, e(X_i^T \theta^*_m) \right) \in IR^{p+m}.
\]

**Lemma A.5.** We have
\[
\|s_i,m(v^*_m)\| \leq (C_{f\|} + 1)\sqrt{26} sX \|\psi'\|_{\infty} + \sqrt{13} \|\psi\|_{\infty} \sqrt{m},
\]
and for any $v, v' \in \mathcal{Y}_0(x)$ with $m^{2\gamma/5}x / \sqrt{m} \to 0$
\[
\|s_{i,\infty}(v) - s_{i,\infty}(v')\| \leq \sqrt{26} (sX \|\psi'\|_{\infty}m^{3/2} + 2(C_{f\|} + 1) \sqrt{m} \|\psi''\|_{\infty} sX
\]
\[
+ 2 \|\psi''\|_{\infty} sX m^{3/2} + \|\psi'\|_{\infty} C_{\eta_m} \sqrt{2L \nabla \phi}) \frac{\|D_{m/2}^1(v - v')\|}{\sqrt{nC_D}}.
\]

**Proof.** Note
\[
\|s_i,m(v^*_m)\| = \|(f^*_\eta(X_i^T \theta^*_m)\nabla \phi(\theta^*_m)^T X_i, e(X_i^T \theta^*_m)\|
\]
\[
\leq \|(f^*_\eta(X_i^T \theta^*_m))\|\|X_i\| + \|e(X_i^T \theta^*_m)\|.
\]
Now because of the wavelet structure and the choice $m = j_m 12 + 2^{j_m} - 1$ we have for each $j = 0, \ldots, j_m - 1$ that
\[
|M(j)| \overset{\text{def}}{=} \left| \left\{ k \in \{j 12 + 2^j, \ldots, (j + 1)12 + 2^{j+1} - 1 \} : |e_k(X_i^T \theta^*_m)| \neq 0 \right\} \right| \leq 13.
This implies

\[ \| e(\mathbf{X}_i^T \theta_m^*) \| = \left( \sum_{k=0}^{m-1} |e_k(\mathbf{X}_i^T \theta_m^*)|^2 \right)^{1/2} = \left( \sum_{j=0}^{j_m-1} \sum_{k \in M(j)} |e_k(\mathbf{X}_i^T \theta_m^*)|^2 \right)^{1/2} \]  

(A.3)

\[ \leq \sqrt{13} \| \psi \|_\infty \left( \sum_{j=0}^{j_m-1} 2^j \right)^{1/2} = \sqrt{13} \| \psi \|_\infty 2^{j_m/2} \leq \sqrt{13} \| \psi \|_\infty \sqrt{m}. \]

Using assumption (Cond\(f^*\)), that \(|M(j)| \leq 13\) and \(k = 12j_k + 2^{j_k} + r_k\) with \(r_k \in \{0, \ldots, 2^{j_k} + 1\}\) and \(j_k \in \mathbb{N}\) we find as \(\alpha > 2\)

\[ |f'_{\eta_m^*}(\mathbf{X}_i^T \theta_m^*)| \leq \sum_{j=0}^{j_m-1} \sum_{k \in M(j)} |\eta_{mk}^*| |e_k(\mathbf{X}_i^T \theta_m^*)| \]

\[ \leq \sqrt{13} \| \psi' \|_\infty \left( \sum_{j=0}^{j_m-1} 2^j \right)^{1/2} \left( \sum_{j=0}^{j_m-1} 2^{-4j} 2^{2j} \right)^{1/2} \]

\[ \leq \sqrt{13} \| \psi' \|_\infty \left( \sum_{k=0}^{m-1} |\eta_{mk}^*|^2 k^4 \right)^{1/2} \left( \sum_{j=0}^{j_m-1} 2^{-j} \right)^{1/2} \]

\[ \leq \sqrt{26} \| \psi' \|_\infty C_{\| \eta_m^* \|}, \]

where with lemma 3.2 and \(m \in \mathbb{N}\) large enough (\(m^5/n \to 0\) and \(x^* \cong m\))

\[ C_{\| \eta_m^* \|} \leq \left( \sum_{k=1}^{m-1} |\eta_{mk}^*|^2 k^4 \right)^{1/2} \leq \left( \sum_{k=0}^{m-1} |\eta_{mk}^*|^2 k^4 \right)^{1/2} + \left( \sum_{k=0}^{m-1} |\eta_{mk}^* - \eta_{mk}^*|^2 k^4 \right)^{1/2} \]

\[ \leq C_{\| f \|} + m^2 \| (\eta_m^* - \Pi_m f^*) \| \]

\[ \leq C_{\| f \|} + \frac{m^2 x^*}{\sqrt{ncD}} \leq C_{\| \eta^* \|} + 1, \]

such that

\[ \|(f'_\infty(\eta_m^*) (\mathbf{X}_i^T \theta_m^*) \nabla \Phi_{\theta_m^*} \mathbf{X}_i, e(\mathbf{X}_i^T \theta_m^*))\| \]

(A.4)

\[ \leq \|f'_\infty(\eta_m^*) (\mathbf{X}_i^T \theta_m^*) \nabla \Phi_{\theta_m^*} \mathbf{X}_i \| + \|e(\mathbf{X}_i^T \theta_m^*)\| \]

\[ \leq (C_{\| f \|} + 1) \sqrt{26} x \| \psi' \|_\infty + \sqrt{13} \| \psi \|_\infty \sqrt{m}. \]

For the second claim we use that for each \(j = 1, \ldots, j_m - 1\)

\[ |N(j)| \overset{\text{def}}{=} \left\{ k \in \{j12 + 2^i, \ldots, (j + 1)12 + 2^{j+1} - 1\} : |e_k(\mathbf{X}_i^T \theta') - e_k(\mathbf{X}_i^T \theta)| \vee |e'_k(\mathbf{X}_i^T \theta) - e'_k(\mathbf{X}_i^T \theta)| > 0 \right\} \leq 26. \]
Further always have that

\[ |e'_k(X_i^\top \theta') - e'_k(X_i^\top \theta)| \leq 2^{j_k} \psi'' \|\|_{\infty} s_X \|\|_{\infty} \theta - \theta'|. \]

This gives again using that \( \alpha > 2 \)

\[
|f'_\infty(\eta^*_m) (\theta^\top X_i) - f'_\infty(\eta^*_m) (X_i^\top \theta')|
\leq \left( \sum_{j=0}^{m-1} \sum_{k \in N(j)} \eta^*_{mk} 2^{5j/2} \right) \|\|_{\theta - \theta'}\|\|_{\psi''} \|\|_{\infty} s_X
\leq \sqrt{26} \left( \sum_{k=0}^{m-1} \eta^*_{mk}^2 k^{2\alpha} \right)^{1/2} \left( \sum_{j=0}^{m-1} 2^{(5-2\alpha)j} \right)^{1/2} \|\|_{\theta - \theta'}\|\|_{\psi''} \|\|_{\infty} s_X,
\]

and with the same arguments

\[
\|e(X_i^\top \theta) - e(X_i^\top \theta')\| \leq \left( \sum_{k=1}^{m} |e_k(X_i^\top \theta) - e_k(X_i^\top \theta')|^2 \right)^{1/2}
\leq \sqrt{26} \left( \sum_{j=0}^{m-1} 2^{3j} \right)^{1/2} \|\|_{\theta - \theta'}\|\|_{\psi''} \|\|_{\infty} s_X,
\]

and

\[
\|f'_{\eta - \eta'} (\theta^\top X_i) \nabla \Phi_{\theta}^\top X_i\| \leq s_X \sum_{k=1}^{m} |\eta_k - \eta'_{mk,k}| \psi' \left( \sum_{j=0}^{m-1} 2^{3j} \right)^{1/2}
\leq \sqrt{26} \|\| \eta - \eta'\|\|_{s_X} \|\|_{\psi''} \|\|_{\infty} m^{3/2}.
\]
Finally similar to (A.4) we have

\[ |f'_{\eta'}(X_i^T \theta')| \leq \sum_{j=0}^{j_{m-1}} \sum_{k \in M(j)} |\eta'_k| |e'_k(X_i^T \theta')| \]

\[ \leq \sqrt{13} \|\psi'\|_{\infty} \left( \sum_{j=0}^{j_{m-1}} \sum_{k \in M(j)} |\eta'_k|^2 2^{4j} \right)^{1/2} \left( \sum_{j=0}^{j_{m-1}} 2^{-4j} 2^j \right)^{1/2} \]

\[ \leq \sqrt{13} \|\psi'\|_{\infty} \left( \sum_{k=0}^{m-1} |\eta'_k|^2 k^4 \right)^{1/2} \left( \sum_{j=0}^{j_{m-1}} 2^{-j} \right)^{1/2} \]

\[ \leq \sqrt{26} \|\psi'\|_{\infty} (C_{\|\eta'\|} + 1), \]

where since \( \psi' \in Y_c(x) \) and \( m \in \mathbb{N} \) large enough

\[ \left( \sum_{k=1}^{m-1} |\eta'_k|^2 k^4 \right)^{1/2} \leq \left( \sum_{k=0}^{m-1} |\eta'_k|^2 k^4 \right)^{1/2} + \left( \sum_{k=0}^{m-1} |\eta'_k - \eta'_k|^2 k^4 \right)^{1/2} \]

\[ \leq C_{\|f\|} + m^2 \left( \|\eta' - \eta'_m\| + \|\eta'_m - \Pi_m \eta'\| \right) \]

\[ \leq C_{\|f\|} + \frac{m^2 (x + x^*)}{\sqrt{nC_D}} \leq C_{\|\eta'\|} + 1, \]

such that

\[ \|f'_\infty(\eta')(X_i^T \theta') \nabla \Phi_{\theta X_i} e(X_i^T \theta')\| \]

\[ \leq \|f'_\infty(\eta')(X_i^T \theta') \nabla \Phi_{\theta X_i} e(X_i^T \theta')\| + \|e(X_i^T \theta')\| \]

\[ \leq (C_{\|f\|} + 1)\sqrt{26sX} \|\psi'\|_{\infty} + \sqrt{13} \|\psi'\|_{\infty} \sqrt{m}. \]

We get combining (A.8), (A.5), (A.7) and (A.6)

\[ \|s_i(\psi) - s_i(\psi')\| \]

\[ = \|f'_\eta(\theta^T X_i) \nabla \Phi_{\theta X_i} \|
\]

\[ + \left[ f'_\infty(\eta') \nabla \Phi_{\theta X_i} e(X_i^T \theta') - f'_\infty(\eta') \nabla \Phi_{\theta X_i} e(X_i^T \theta') \right] \nabla \Phi_{\theta X_i} e(X_i^T \theta') \]

\[ \leq \sqrt{26} \|\eta - \eta'\| sX \|\psi'\|_{\infty} m^{3/2} + \sqrt{26}(C_{\|f\|} + 1)\sqrt{m} \|\theta - \theta'\| \|\psi''\|_{\infty} sX \]

\[ + \sqrt{26} m^{3/2} \|\theta - \theta'\| \|\psi'\|_{\infty} sX \]

\[ + \|\psi'\|_{\infty} sX m^{3/2} + \|\psi'\|_{\infty} C_{\|\eta'\|} \sqrt{26} L_{\psi'\theta} \|\theta - \theta'\| \]

\[ \leq \sqrt{26} sX \|\psi'\|_{\infty} m^{3/2} + 2(C_{\|f\|} + 1)\sqrt{m} \|\psi''\|_{\infty} sX \]

\[ + \|\psi'\|_{\infty} sX m^{3/2} + \|\psi'\|_{\infty} C_{\|\eta'\|} \sqrt{26} L_{\psi'\theta} \]
where we used Lemma A.4 in the last step to find that
\[
\|\theta - \theta'\| \vee \|\eta - \eta'\| \leq \sqrt{\|\theta - \theta'\|^2 + \|\eta - \eta'\|^2} \leq \|v - v'\|
\]
\[
\leq \frac{\|D_{m}^{1/2}(v - v')\|}{\sqrt{nC_2}}.
\]

The next auxiliary Lemma relies on a non-commutative Bernstein inequality; see Koltchinskii (2012):

**Lemma A.6.** Suppose that \(s_{i,m} \in \mathbb{R}^m\) are iid random vectors. Define
\[
S_n^* := \frac{1}{n} \sum_{i=1}^{n} s_{i,m}s_{i,m}^\top - IE[s_{1,m}s_{1,m}^\top] =: M_i,
\]
and \(B^2 := IE[\|s_{1,m}\|^4]\). Assume that \(\|s_{i,m}s_{i,m}^\top\| = \|M_i\| \leq U \in \mathbb{R}\) then it holds
\[
IP(\|S_n^*\| > n^{-1}t) \leq 2m \exp\left\{- \frac{t^2}{4nB^2 + 2Ut/3}\right\}.
\]

**Proof.** This lemma is an immediate consequence of the non-commutative Bernstein inequality in Koltchinskii (2012). We only have to note that
\[
\sum_{i=1}^{n} IE[M_i^2] \leq 2n IE[\|s_{1,m}\|^4] = 2nB^2.
\]

**Lemma A.7.** We have with \(t = c_M \sqrt{s_{m} x + \log(2m)}\) where \(c_M = 26\left(C_2f + 1\right)\sqrt{2s_{X} \psi'_{\infty} + \psi_{\infty}^2}\) and where \(x \leq 9n/2 - \log(2m)\) that
\[
IP(\|S_n\| \geq n^{-1}t) \leq e^{-x},
\]
where
\[
S_n = \frac{1}{n} \sum_{i=1}^{n} s_{i,m}(v_{m}^*)s_{i,m}(v_{m}^*)^\top - \frac{1}{n} \eta_{m}^2(v_{m}^*).
\]

**Proof.** We want to employ lemma A.6. We estimate using Lemma A.5
\[
\|s_{i,m}(v_{m}^*)s_{i,m}(v_{m}^*)^\top\| \leq 26\left(C_2f + 1\right)\sqrt{2s_{X} \psi'_{\infty} + \psi_{\infty}^2} m =: C_M m,
\]
where the same bound holds for the expected value such that \(\|s_{i,m}s_{i,m}^\top - IE[s_{1,m}s_{1,m}^\top]\| =: \|M_i\| \leq C_M m\). Further
\[
IE[\|s_{i,m}(v_{m}^*)\|^4] \leq C_M^2 m^2.
\]
Lemma A.8. We have with $t = C_M \| D_m^{1/2}(\mathbf{v} - \mathbf{v'})\|^2 \sqrt{5/nm^3} \left( x + \log(2m) \right)^{1/2}$ and $x \leq 9n/2 - \log(2m)$ this gives

$$\mathbb{P}(\| \mathbf{s}_n \| \geq n^{-1} t) \leq e^{-x}.$$  

\[ \square \]

**Lemma A.8.** We have with $t = C_M^{2} \| D_m^{1/2}(\mathbf{v} - \mathbf{v'})\|^2 \sqrt{5/nm^3} \left( x + \log(2m) \right)^{1/2}$ and $x \leq 9n/2 - \log(2m)$

$$\mathbb{P}(\| \mathbf{s}_n \| \geq n^{-1} t) \leq e^{-x},$$

where with $\mathbf{v} \in \mathcal{T}_a(\mathbf{r})$

$$\mathbf{s}_n = \frac{1}{n} \sum_{i=1}^{n} \left( \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \right) \left( \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \right)^T$$

$$- \mathbb{E}\left( \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \right) \left( \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \right)^T$$

$$C_M = \sqrt{26} \left( s \| \psi' \|_\infty + 3(C\| \mathbf{f} \| + 1)\| \psi'' \|_\infty s \mathbf{x} \right)$$

$$+ 3\| \psi' \|_\infty s \mathbf{x} + \| \psi' \|_\infty C\| \eta' \| \sqrt{2L \nabla \phi} \right) \frac{2}{c_D}.$$  

**Proof.** We estimate using Lemma A.5

$$\| (\mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v})) (\mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}))^T \|$$

$$\leq \| \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \|^2$$

$$\leq 26 \left( s \| \psi' \|_\infty + 3(C\| \mathbf{f} \| + 1)\| \psi'' \|_\infty s \mathbf{x} \right)$$

$$+ 3\| \psi' \|_\infty s \mathbf{x} + \| \psi' \|_\infty C\| \eta' \| \sqrt{2L \nabla \phi} \right) \frac{4\| D_m^{1/2}(\mathbf{v} - \mathbf{v'})\|^2 m^3}{nc_D^2}$$

$$=: C_M^2 \| D_m^{1/2}(\mathbf{v} - \mathbf{v'})\|^2 m^3$$

where the same bound holds for the expected value such that

$$\| (\mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v})) (\mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}))^T - \mathbb{E}\left( \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \right) \left( \mathbf{s}_{i,m}(\mathbf{v'}) - \mathbf{s}_{i,m}(\mathbf{v}) \right)^T \|$$

$$=: \| \mathbf{M}_i \| \leq 2C_M^2 t m^3 \| D_m^{1/2}(\mathbf{v} - \mathbf{v'})\|^2,$$
for some $C_M \in \mathcal{R}$. With the same estimates we obtain

$$
\mathbb{E}[\|\hat{s}_{i,m}(v') - s_{i,m}(v)\|^4] \leq C_M^4 m^6 \frac{\|D_m^{1/2}(v - v')\|^4}{n^2}.
$$

Plugging these bounds into lemma A.6 we get

$$
\mathbb{P}(\|S_n\| \geq n^{-1}t) 
\leq 2m \exp\left\{-\frac{t^2}{4\|D_m^{1/2}(v - v')\|^4 C_M^4 n^{-1}m^6 + 2\|D_m^{1/2}(v - v')\|^2 C_M^2 m^3 n^{-1/3}}\right\}.
$$

Setting $t = C_M^2 \|D_m^{1/2}(v - v')\|^2 \sqrt{8/3n} (x + \log(2m))^{1/2}$ and $x \leq 9n/2 - \log(2m)$ this yields

$$
\mathbb{P}(\|S_n\| \geq n^{-1}t) \leq e^{-x}.
$$

\[\square\]

A.3.1 Conditions ($ED_0$), ($E\tau$) and ($ED_{1,m}$)

Lemma A.9. On a set of dominating probability we have ($ED_0$) with

$$
g = \sqrt{n}\sigma^{-1} c_{ED} \tilde{g}( (C_{||n^*||} + 1) \sqrt{26s} \|\psi'\|_{\infty} + \sqrt{13}\|\psi\|_{\infty} \sqrt{m})^{-1},
$$

$$
\nu_{r,m}^2 = \tilde{\nu}_r^2 \left( \|V_m^{-1}(v^s)\|_{\infty}^{1/2} \|V_{uv}^{-1}(v^s)\|_{\infty} + 1 \right),
$$

and ($E\tau$) with

$$
g(\tau) = \sqrt{n} c_{ED} \tilde{g} (C_{\|f\|} m^{-3/2},
$$

$$
\nu_{r,m}^2 = \tilde{\nu}_r^2 \left( c_{ED}(\tau^s)^2 + C_{(E\tau)} m^3/\sqrt{n} \right),
$$

where $c_{ED}(\tau^s)^2 = \sup_{v \in \mathcal{R}_0(\sqrt{n}\tau^s)} \|V_m^2(v)\|^{-1/2} \|V_{uv}^{1/2}(v^s)\|$ with some $\tau^s > 0$ and where $C_{(E\tau)} > 0$ is independent of $n, m, x$ can be bounded by

$$
c_{(E\tau)} \leq \sqrt{26} \left( s \|\psi'\|_{\infty} + 3(C_{\|f\|} + 1) \|\psi'\|_{\infty} s \|x\| \right. 
\left. + 3\|\psi'\|_{\infty} s \|x\| C_{\eta_m} \sqrt{2L_{\psi'}} \right) \frac{2}{c_{ED}} + (C_{\|f\|} + 1) \sqrt{26} s \|\psi'\|_{\infty} + \sqrt{13}\|\psi\|_{\infty}.
$$

Proof. Lemma A.4 gives with $\tilde{\gamma} = V_{uv}^{1/2} \gamma/\|V_{uv}^{1/2}\|

$$
\frac{\langle \nabla \xi(v_m^s), \gamma \rangle_{R^p}}{\|D_m^\gamma \|} = \langle \tilde{\gamma}^T D_m^{-1} A(v_m^s), \xi \rangle_{R^o}.
$$
Consequently - using Lemma A.5 - we get with $\mu \leq \sqrt{n\sigma^{-1}}c_D\bar{g}\left((C_{\|\eta^*\|}+1)\sqrt{26sX}\|\psi\|_\infty+\sqrt{13}\|\psi\|_{\infty}\sqrt{m}\right)^{-1}$ and assumption (Cond.)

$$
\sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \left( \frac{\langle \nabla \zeta(v_m^s), \gamma \rangle}{\|D_m(v^s)\gamma\|} \right) \right\}
\leq \sum_{i=1}^{n} \sup_{\gamma \in \mathbb{R}^p, \|\gamma\|=1} \log \mathbb{E} \exp \left\{ \mu(\tilde{\gamma}, D_m^{-1}(v^s)\xi_{i,m}(v_m^s))\varepsilon_i \right\}
\leq \nu^2 \mu^2 \gamma^\top D_m^{-1}(v^s) \left( \sum_{i=1}^{n} \xi_{i,m}(v_m^s)\xi_{i,m}(v_m^s)\right) D_m^{-1}(v^s) \tilde{\gamma}
= \nu^2 \mu^2 \gamma^\top D_m^{-1}(v^s) D_m(v_m^s) D_m^{-1}(v^s) \tilde{\gamma}
+ \nu^2 \mu^2 \gamma^\top D_m^{-1}(v^s) nS_n D_m^{-1}(v^s) \tilde{\gamma}
\leq \nu^2 \mu^2 \|D_m^{-1}(v^s)\|_2 D_m^{1/2}(v_m^s) \|^2 + \nu^2 \mu^2 \kappa_n,
$$

(A.9)

where

$$
\kappa_n = \tilde{\gamma}^\top (n^{-1}D_m)^{-1/2} S_n (n^{-1}D_m)^{-1/2} \tilde{\gamma},
$$

$$
S_n = \frac{1}{n} \sum_{i=1}^{n} \xi_{i,m}(v_m^s)\xi_{i,m}(v_m^s)^\top - \frac{1}{n} D_m(v_m^s).
$$

With Lemma A.7 we infer that with $t = c_M \sqrt{5nm} \left(x + \log(2m)\right)^{1/2}$ where $c_M = \left((C_{\|f\|}+1)\sqrt{26sX}\|\psi\|_\infty + \sqrt{13}\|\psi\|_{\infty}\right)^2$ the set $\{\|S_n\| \leq n^{-1}\}$ is of dominating probability if $x \leq 9n/2 - \log(2m)$. Consequently with probability greater $1 - e^{-x}$ we find that for $n \in \mathbb{N}$ large enough

$$
\kappa_n \leq \frac{c_M \sqrt{5m} \left(x + \log(2m)\right)^{1/2}}{\sqrt{n\sigma^2 c_D^2}} \leq 1.
$$

Thus when $m \left(x + \log(2m)\right)/\sqrt{n} \leq \sqrt{5c_M}/(\sigma^2 c_D^2)$ we get $(ED_0)$ with probability greater $1 - e^{-x}$ and

$$
g = \sqrt{n}c_D\bar{g}\left((C_{\|\eta^*\|}+1)\sqrt{26sX}\|\psi\|_\infty + \sqrt{13}\|\psi\|_{\infty}\sqrt{m}\right)^{-1},
\nu^2 = \nu^2 \left(\|D_m^{-1}(v^s)\|_2 + 1\right).
$$

Concerning $(ER)$ we observe that

$$
\|\xi_{i,m}(v)\| \leq \|\xi_{i,m}(v_m^s)\| + \|\xi_{i,\infty}(v) - \xi_{i,\infty}(v_m^s)\|. 
$$
Due to Lemma A.5 on $\Upsilon_0(r)$

$$\|s_{i,m}(v^*_m)\| \leq (C_{\eta^*} + 1)\sqrt{26s_X}\|\psi'\|_{\infty} + \sqrt{13}\|\psi\|_{\infty}\sqrt{m} := C_{1,M}\sqrt{m},$$

and

$$\|s_{i,\infty}(v) - s_{i,\infty}(v^*_m)\| \leq \sqrt{26}(s_X\|\psi'\|_{\infty}m^{3/2} + 3(C_{f\|f\|} + 1)\sqrt{m}\|\psi''\|_{\infty}s_X$$

$$+ 3\|\psi''\|_{\infty}s_Xm^{3/2} + \|\psi'\|_{\infty}C_{\eta_m}\sqrt{2L\nabla \Phi})\frac{2r}{\sqrt{nCD}}$$

$$:= C_{2,M}r_m^{3/2}/\sqrt{n}.$$
Proof. We get with Lemma A.5 and with Lemma A.4

\[
\langle D_m^{-1} \gamma, \nabla \zeta(v) - \nabla \zeta(v^*_m) \rangle \\
\leq \frac{\sqrt{26}}{\sqrt{nC_D}} \sum_{i=1}^{n} \varepsilon_i \left( s_X \|\psi'\|_\infty m^{3/2} + 3(C\|f\| + 1)\sqrt{m}\|\psi''\|_\infty s_X \right. \\
\left. + 3\|\psi'\|_\infty s_X m^{3/2} + \|\psi'\|_\infty C\|\eta_n\| \sqrt{2L}\psi \right) \frac{2\pi}{\sqrt{nC_D}}.
\]

We get with,

\[
\mu \leq g \defeq \sqrt{nC_D} m^{-3/2} 26^{-1/2} \left( s_X \|\psi'\|_\infty + 3(C\|f\| + 1)\|\psi''\|_\infty s_X \right. \\
\left. + 3\|\psi'\|_\infty s_X + \|\psi'\|_\infty C\|\eta_n\| \sqrt{2L}\psi \right)^{-1}
\]

\[
\omega \defeq \frac{2}{\sqrt{nC_D}},
\]

and the same calculations as in (A.9) with some \(v, v' \in \mathcal{Y}_o(x)\), \(\gamma \in \mathbb{R}^p\) and \(\|\gamma\| = 1\)

\[
\log \mathbb{E}[\exp \left\{ \mu \frac{\gamma^\top D_m^{-1} (\nabla \zeta(v) - \nabla \zeta(v'))}{\omega\|D_m^{1/2}(v - v')\|} \right\}]
\]

\[
\leq \sum_{i=1}^{n} \log \mathbb{E}[\exp \left\{ \mu \varepsilon_i \frac{\gamma^\top D_m^{-1} (s_{i,m}(v) - s_{i,m}(v'))}{\omega\|D_m^{1/2}(v - v')\|} \right\}]
\]

\[
\leq \mu^2 \nu^2 (\omega\|D_m^{1/2}(v - v')\||)^{-2}
\]

\[
\gamma^\top D_m^{-1} \left( \sum_{i=1}^{n} (s_{i,m}(v') - s_{i,m}(v))(s_{i,m}(v') - s_{i,m}(v))^\top \right) D_m^{-1} \gamma^\top.
\]

We estimate

\[
\gamma^\top D_m^{-1} \left( \sum_{i=1}^{n} (s_{i,m}(v') - s_{i,m}(v))(s_{i,m}(v') - s_{i,m}(v))^\top \right) D_m^{-1} \gamma^\top \leq \|D_m^{-1} n \mathbb{E} \left[ (s_{i,m}(v') - s_{i,m}(v))(s_{i,m}(v') - s_{i,m}(v))^\top \right] D_m^{-1}\| + \varepsilon_n
\]

\[
\leq \mathbb{E} \| (n^{-1} D_m)^{-1/2} (s_{i,m}(v') - s_{i,m}(v))^2 + \varepsilon_n
\]

\[
def \varphi(v, v') + \varepsilon_n,
\]

where

\[
\varepsilon_n = \| (n^{-1} D_m)^{-1/2} S_n (n^{-1} D_m)^{-1/2} \|
\]

\[
S_n = \frac{1}{n} \sum_{i=1}^{n} (s_{i,m}(v') - s_{i,m}(v))(s_{i,m}(v') - s_{i,m}(v))^\top
\]

\[
- \mathbb{E}(s_{i,m}(v') - s_{i,m}(v))(s_{i,m}(v') - s_{i,m}(v))^\top.
\]
 Finite sample analysis of profile M-estimation in the Single Index model

To control $\kappa_n > 0$ we apply lemma A.8 and we infer that with $t = C_M^2 \| D_m^{1/2} (\mathbf{v} - \mathbf{v}') \|^2 \sqrt{5/nm^3} \left( x + \log(2m) \right)^{1/2}$ and $x \leq 9n/2 - \log(2m)$ the set $\{ \| S_n \| \leq n^{-1} t \}$ is of dominating probability and on this set we find

$$\kappa_n \leq \frac{C_M^2 \| D_m^{1/2} (\mathbf{v} - \mathbf{v}') \|^2 \left( x + \log(2m) \right)^{1/2}}{nc_D^2} \sqrt{\frac{5}{n}} \left( x + \log(2m) \right)^{1/2} \frac{1}{\sqrt{\overline{D}}} C \| \eta_m \| \sqrt{2L \nabla \Phi} \right)^2 \frac{\| D_m^{1/2} (\mathbf{v} - \mathbf{v}') \|}{nc_D^4}.$$

Further we have with Lemma A.5

$$\varphi(\mathbf{v}, \mathbf{v}') \leq 26 \left( sX \| \psi' \|_\infty m^{3/2} + 2(C \| f \| + 1) \sqrt{m} \| \psi'' \|_\infty sX \right)$$

$$+ 2 \| \psi' \|_\infty sX m^{3/2} + \| \psi' \|_\infty C \| \eta_m \| \sqrt{2L \nabla \Phi} \right)^2 \frac{\| D_m^{1/2} (\mathbf{v} - \mathbf{v}') \|}{nc_D^4} \frac{1}{c_D^2}.$$

Together this gives $(\mathcal{E D}_1)$ with

$$\nu^2_{1,m} = \nu^2 \frac{C_M^2 m^3}{n} \left( \frac{1}{c_D^2} + \frac{\sqrt{5}}{n} \left( x + \log(2m) \right)^{1/2} \right).$$

\[ \square \]

A.3.2 Condition $(L_0)$

**Lemma A.11.** The condition $(L_0)$ is satisfied where with $\mathbf{r}^* > 0$ from equation (3.2)

$$\delta(\mathbf{r}) = \frac{\left( C_{\delta,1} m^2 + C_{\delta,2} \left( m^{3/2} \sqrt{\frac{m^3}{\sqrt{n}}} \right) \right)}{c_D \sqrt{n}} \left[ \mathbf{r} + \mathbf{r}^* \right],$$

where $C_{\delta,1}, C_{\delta,2} > 0$ are polynomials of $\| \psi \|_\infty, \| \psi' \|_\infty, \| \psi'' \|_\infty, C \| f \|, L \nabla \Phi, sX$.

**Proof.** We will show that $\frac{1}{n} \| D_m^2 (\mathbf{v}) - D_m^2 (\mathbf{v}^*) \| \leq c_D^2 \delta(\mathbf{r})$, which will give the claim due to

$$\| I_p^* - D_m^{-1} \nabla^2 \mathbf{E} [L(\mathbf{v})] D_m^{-1} \| \leq \frac{1}{nc_D^2} \| D_m^2 (\mathbf{v}) - D_m^2 (\mathbf{v}^*) \|. $$
\[
-\nabla_p^2 \mathbb{E}[\mathcal{L}_\lambda(v)] \overset{\text{def}}{=} \mathcal{D}_m^2(v) = nd_m^2(v) + nr_m^2(v),
\]

\[
n d_m^2 = n \begin{pmatrix}
 a_m(v) \\
 h_m^2(v)
\end{pmatrix} = \begin{pmatrix}
 D(v) \\
 A_m(v) \\
 H_m^2(v)
\end{pmatrix},
\]

\[
r_m^2(v) = \mathbb{E} \left[ (f_\eta(X^\top \theta) - f_\eta^*(X^\top \theta^*)) \begin{pmatrix}
 v_\theta^2(v) \\
 b_m(v)
\end{pmatrix} \right],
\]

\[
v_\theta^2(v) = 2f''_\eta(X^\top \theta) \nabla \Phi^\top \theta X^\top \nabla \Phi + |f'_\eta(X^\top \theta)|^2 \nabla^2 \Phi^\top \theta \nabla X,
\]

\[
b_m(v) = \nabla \Phi \theta^\top e'(X^\top \theta),
\]

such that

\[
\frac{1}{n} \| \mathcal{D}_m^2(v) - \mathcal{D}_m^2(v^*) \| \leq \frac{1}{n} \left( \| D^2(v) - D^2(v^*) \| + 2 \| A_m(v) - A_m(v^*) \| + \| H_m^2(v) - H_m^2(v^*) \| + \| r_m^2(v) \| \right),
\]

so that we can calculate separately

\[
\frac{1}{n} \| D^2(v) - D^2(v^*) \| 
\leq \mathbb{E}[\| X \|^2 \left\{ |(f'_\eta)^2 - (f'_{\eta^*})^2| X^\top \theta | + |(f'_{\eta^*})^2 - (f'_\eta)^2| X^\top \theta^* | + 2 |(f'_{\eta^*})^2 X^\top \theta^* | \| \nabla \Phi - \nabla \Phi^* \| \right\}].
\]

Using (A.4) we find

\[
|(f'_{\eta^*})^2(X^\top \theta^*) | \| \nabla \Phi - \nabla \Phi^* \| \leq \| \psi' \|_\infty (C_{\| f \|} + 1) \sqrt{2L \nabla \Phi} \| \theta - \theta^* \|.
\]

Further we have

\[
|(f'_\eta - f'_{\eta^*})(X^\top \theta)| \leq \| \eta - \eta^* \| \left( \sum_{k=0}^m |e_k'(X^\top \theta)|^2 \right)^{1/2} \overset{(A.10)}{=} \| \eta - \eta^* \| \| \psi' \|_\infty m^{3/2},
\]
This gives using (A.4) and (A.10)

\[
\begin{aligned}
\left| \left( f'\eta - f'\eta^* \right)^2 (X^T \theta) \right| \\
&\leq \left( \left| f'\eta (X^T \theta) \right| + \left| f'\eta^* (X^T \theta) \right| \right) \left| f'\eta - f'\eta^* \right| (X^T \theta) \\
&\leq \left( \left| (f'\eta - f'\eta^*) (X^T \theta) \right| + 2 \left| f'\eta^* (X^T \theta) \right| \right) \left| (f'\eta - f'\eta^*) (X^T \theta) \right| \\
&\leq \left( 2 \|\psi'\|_\infty (C\|f\| + 1)^2 \sqrt{2} \sqrt{m} \|\psi''\|_\infty sX \|\theta - \theta^*\| \right) \|\eta - \eta^*\|.
\end{aligned}
\]

Finally we derive with (A.5) and (A.4)

\[
\begin{aligned}
\left| \left( f'\eta^* \right)^2 (X^T \theta) - \left( f'\eta^* \right)^2 (X^T \theta^*) \right| \\
&\leq \left( \left| f'\eta^* (X^T \theta^*) \right| + \left| f'\eta^* (X^T \theta^*) \right| \right) \left| f'\eta^* (X^T \theta) - f'\eta^* (X^T \theta^*) \right| \\
&\leq 4 \sqrt{2} \|\psi'\|_\infty (C\|f\| + 1)^2 \sqrt{m} \|\psi''\|_\infty sX \|\theta - \theta^*\|.
\end{aligned}
\]

Collecting everything yields

\[
\begin{aligned}
\frac{1}{n} \|D^2(\nu) - D^2(\nu^*)\| &\leq s_X^2 \left\{ \left( 2 \|\psi'\|_\infty (C\|f\| + 1)^2 \sqrt{2} \sqrt{m} \|\psi''\|_\infty sX \|\theta - \theta^*\| \right) \|\eta - \eta^*\| \\
&\leq \mathcal{C}_1 \left( m^{3/2} \sqrt{\frac{\Sigma m^3}{\sqrt{n}}} \right) \|\nu - \nu^*\|.
\end{aligned}
\]

With the same arguments, additionally using (A.3) we find similar bounds for the second and the third summand via adapting the constant \( C(\alpha) > 0 \)

\[
\begin{aligned}
\frac{1}{n} \|A_m(\nu) - A_m(\nu^*)\| &\leq \mathcal{C}_2 \left( m^{3/2} \sqrt{\frac{\Sigma m^3}{\sqrt{n}}} \right) \|\nu - \nu^*\|, \\
\frac{1}{n} \|H_m^2(\nu) - H_m^2(\nu^*)\| &\leq \mathcal{C}_3 \left( m^{3/2} \sqrt{\frac{\Sigma m^3}{\sqrt{n}}} \right) \|\nu - \nu^*\|. \quad (A.11)
\end{aligned}
\]
Finally we estimate the fourth term. First note that again using the wavelet structure

\[ |f''_\eta(X^\top \theta)| \leq |f''_{\eta-\eta^*}(X^\top \theta)| + |f''_{\eta^*}(X^\top \theta)| \]

\[ \leq \sqrt{26}\|\phi''\|_\infty \left( \sum_{k=0}^{m} (\eta_k - \eta_k^*)^2 \right)^{1/2} \left( \sum_{j=0}^{j_{m-1}} 2^{5j} \right)^{1/2} \]

\[ + \sqrt{26}\|\phi''\|_\infty \left( \sum_{j=0}^{j_{m-1}} \eta_k^* 2^{2\alpha} \right)^{1/2} \left( \sum_{j=0}^{j_{m-1}} 2^{(5-2\alpha)j} \right)^{1/2} \]

\[ \leq \sqrt{26}\|\phi''\|_\infty \|D_{m}^{1/2}(v - v^*)\| \leq \sqrt{26}\|\phi''\|_\infty C_{\eta^*} \sqrt{2m^{1/2}}. \]

Further using (A.4) we have for any \( \varphi \in \mathbb{R}^{p-1} \) with \( \|\phi\| = 1 \)

\[ ||f'_\eta(X^\top \theta)^2\nabla^2 \phi^\top [X, \varphi, \cdot]||_{\mathbb{R}^p} \leq 26\|\psi\|^2_{\infty} (C_{\eta^*}) + 1)^2 s_X. \]

We also find

\[ |f_f(X^\top \theta) - f_{f_f}(X^\top \theta^*)| \leq \left( \sum_{k=1}^{m} (\eta^*)^2 k^{2\alpha} \right)^{1/2} \left( \sum_{k=1}^{m} |\epsilon'_k(X^\top \theta^*)|^2 k^{-2\alpha} \right)^{1/2} L_{\nabla, \phi} \|X\| \|\theta - \theta^*\| \]

\[ \leq 2\sqrt{26} C_{\eta^*} \sqrt{2L_{\nabla, \phi}} s_X \|\theta - \theta^*\|. \]

This gives for constants \( C_i > 0 \) large enough

\[ \|r_m^2(v)\| \]

\[ \leq \mathbb{E}[|f_{\eta_f}(X^\top \theta) - f_{\eta^*_f}(X^\top \theta^*)||\bar{v}_m^2(v)||] \]

\[ \leq \mathbb{E}\left(|f_{\eta_f}(X^\top \theta) - f_{\eta^*}(X^\top \theta^*)| + |f_{\eta^*}(X^\top \theta)| + |f_f(X^\top \theta) - f_f(X^\top \theta^*)| + \right) \]

\[ \left( \|f''_{\eta^*}(X^\top \theta)\|^2 \|X\|^2 + \sup_{\|\varphi\| = 1} \|f'_\eta(X^\top \theta)^2\nabla^2 \phi^\top [X, \varphi, \cdot]||_{\mathbb{R}^p} + 2s_X|\epsilon'(X^\top \theta)|\right) \]

\[ \leq C_4 m^{3/2} \|v - v^*\| \]

\[ + m^{3/2} C_5 \left( \mathbb{E}[|f_{\eta_f}(X^\top \theta) - f_{\eta^*_f}(X^\top \theta)|^2]^{1/2} + \mathbb{E}[|f_{\eta^*}(X^\top \theta)|^2]^{1/2} \right). \]
We estimate using (A.11) and and a constant $C_6 > 0$ large enough
\[
\mathbb{E}[|f_{\eta^*}(X^T \theta) - f_{\eta^*}(X^T \theta)|^2]^{1/2} = \frac{1}{\sqrt{n}} \|H_m(v)(\eta - \eta^*)\| \\
\leq \frac{1}{\sqrt{n}} \|H_m^2(v) - H_m^2(v^*)\|^{1/2} \|\eta - \eta^*\| + \frac{1}{\sqrt{n}} \|H_m(v^*)(\eta - \eta^*)\| \\
\leq \left( C_3^{1/2} \left( \frac{m^{3/2} \vee \frac{r m^3}{n^{1/2}}}{\sqrt{n}} \right)^{1/2} + 2 \right) \frac{1}{\sqrt{n} \delta D} \|D_m(v - v^*)\| \\
\leq \frac{C_6}{\sqrt{n} \delta D} \|D_m(v - v^*)\|,
\]
and with the same arguments as in the proof of lemma A.2 following equation (A.2)
\[
\mathbb{E}[|f_{\eta^*}(X^T \theta)|^2]^{1/2} \leq \sqrt{n} \left( \sum_{k=m+1}^{\infty} \eta_k^2 \|d_{X^T \theta}\|_\infty + 2 \sqrt{n} \right) \sum_{l>k} \eta_l^* \eta_k^* \mathbb{E}[e_ke_l(X^T \theta^*)] \\
\leq r^*/\sqrt{n}.
\]
Collecting everything this gives using $C_{\alpha-2}(m)^2 \leq \sqrt{m}$
\[
\frac{1}{n} \|D_m^2(v) - D_m^2(v^*)\| \leq \frac{C_7}{\sqrt{n} \delta D} \left( \frac{m^{3/2} \vee \frac{r m^3}{n^{1/2}}}{\sqrt{n}} \right) \|D_m(v - v^*)\| + m^{3/2}C_5 r^*/\sqrt{n}.
\]
We find since $v \in \Upsilon_\delta(\varepsilon)$, with $r^* > 0$ from (3.2) and with lemma A.3 that
\[
\|D_m(v - v^*)\| \leq \|D_m(v - v_m^*)\| + \|D_m(v_m^* - v^*)\| \leq r + r^*;
\]
Such that
\[
\delta(\varepsilon) = \frac{C(C_0) \left( \frac{m^{3/2} \vee \frac{r m^3}{\sqrt{n}}}{\sqrt{n}} \right) [r + r^*]}{c_D \sqrt{n}}.
\]

The claim of Lemma 3.3 follows via combining the above lemmas such that we get the claim on a set which is of dominating probability by the Lemmas A.7 and A.8.

**A.3.3 Condition (L\varepsilon)**

In this Section we will distinguish $\theta \in S_1^{p,+}$ and $\varphi_\theta \in W_\delta$ with $\Phi(\varphi_\theta) = \theta$ from each other.

**Lemma A.12.** Assume the conditions (A). Then for $n \in \mathbb{N}$ large enough there exist $c(Q), c(L\varepsilon), c > 0$ such that with probability $1 - \exp \left\{ -n/m^5 \right\} - \exp \left\{ -nc(Q)/4 \right\}$
\[
- \inf_{v \in \Upsilon_\delta(\varepsilon)^c} \mathbb{E}[\mathcal{L}(v, v_m^*)] > c(L\varepsilon) r^2/2,
\]
as soon as $r^2 \geq C\theta^2$.

Proof. We will proof this claim using an idea by Mendelson (2014). First note that we have with expectation taken conditioned on $(X) = (X_i)_{i=1,...,n} \subset \mathbb{R}^d$

$$-\mathbb{E}[\mathcal{L}(\nu, \nu_m^*)|(X)]$$

$$= \sum_{i=1}^{n} \left[ |f_{\eta_m}(X_i^\top \theta) - f_{\eta}(X_i^\top \theta^*)|^2 - |f_{\eta_m}(X_i^\top \theta^*) - f_{\eta}(X_i^\top \theta^*)|^2 \right]$$

$$\geq \sum_{i=1}^{n} \left[ |f_{\eta_m}(X_i^\top \theta) - f_{\eta}(X_i^\top \theta^*)|^2 - n\mathbb{E}[|f_{\eta_m}(X_i^\top \theta^*) - f_{\eta}(X_i^\top \theta^*)|^2] \right]$$

$$- \sum_{i=1}^{n} \left[ |f_{\eta_m}(X_i^\top \theta^*) - f_{\eta}(X_i^\top \theta^*)|^2 - \mathbb{E}[|f_{\eta_m}(X_i^\top \theta^*) - f_{\eta}(X_i^\top \theta^*)|^2] \right].$$

We will show that with high probability $1-\exp \{-n/m^5\}$ and certain constants $C_m, C_\Sigma > 0$ that may depend on the nuisance dimension $m > 0$

$$n\mathbb{E}[|f_{\eta_m}(X_i^\top \theta^*) - f_{\eta}(X_i^\top \theta^*)|^2] \leq C_m, \quad (A.12)$$

$$\left| (P_n - \mathbb{P})|f_{\eta_m}(X_i^\top \theta^*) - f_{\eta}(X_i^\top \theta^*)|^2 \right| \leq C_\Sigma, \quad (A.13)$$

and that

$$Q(b) \overset{\text{def}}{=} \inf_{(\theta, \eta) \in \mathcal{T}_c} \mathbb{P} \left( |f_{\eta}(X_i^\top \theta) - f_{\eta}(X_i^\top \theta^*)|^2 \geq br^2/n \right) > 0, \quad (A.14)$$

for the right choice of $b > 0$ and $r > r_Q > 0$ large enough.

Lemma A.13. Under (A.14), (A.13) and (A.13) we get

$$\inf_{\nu \in \mathcal{T}_c} -\mathbb{E}[\mathcal{L}(\nu, \nu_m^*)|(X)] \geq \lambda br^2$$

with probability greater $1-\exp \{-n/m^5\} - \exp \{-nQ(2b)^2/4\}$ for $r^2 \geq (C_m + C_\Sigma)/(\lambda b) \vee r_Q^2$ if

$$0 < \lambda \overset{\text{def}}{=} \left( Q(2b) - 2/n + C\sqrt{\log(n)p^*} / n \right) / 4,$$

for a constant $C > 0$ which is a function of $\|\psi\|_{\infty}, \|\psi\|_{\infty}, sX, \text{diam}(Y)$.

Proof. This is a direct consequence of Theorem A.19. It remains to bound using the
Finite sample analysis of profile M-estimation in the Single Index model

proof of Theorem 8.15 of Kosorok (2005)

$$\IE \left[ \sup_{\nu \in \mathcal{T}(r)^c} (P_n - \mathbb{P}) \chi_b(\nu) \right] \leq \IE \left[ \sup_{\nu \in \mathcal{Y}} (P_n - \mathbb{P}) \chi_b(\nu) \right]$$

(A.15)

$$\leq 2C^* \IE \left[ \sqrt{\frac{6(1 + \log N(\delta, \mathcal{F}, L_1(P_n)))}{n}} \right] + \delta,$$

where $N(\delta, \mathcal{F}, L_1(P_n))$ denotes the $\delta$-ball covering number of $\mathcal{F} \overset{\text{def}}{=} \{\chi_b(\nu) : \nu \in \mathcal{Y}\}$ with respect to the norm

$$\|h\|_{L_1(P_n)} = P_n|h(X)| = \frac{1}{n} \sum_{i=1}^{n} |h(X_i)|.$$

The universal constant $C^* > 0$ comes from Lemma 8.2 of Kosorok (2005) ($C^* = K(\exp(x^2) - 1)$). The function $\chi_b : \mathcal{Y} \to \mathbb{R}$ is defined via

$$\chi_b(\nu) \overset{\text{def}}{=} \frac{1}{\|f_\eta(X_i^\top \theta) - f_{\eta^*}(X_i^\top \theta^*)\|^2}.$$

We want to bound the right hand side of (A.15). For this note that

$$\log N(\delta, \mathcal{F}, L_1(P_n)) \leq \log N(\delta/(L(P_n) \vee 1), \mathcal{Y}, \|\cdot\|_2),$$

where

$$L(P_n) = \sup_{\nu, \nu^\circ \in \mathcal{Y}} \frac{\|\chi_b(\nu) - \chi_b(\nu^\circ)\|_{L_1(P_n)}}{\|\nu - \nu^\circ\|_2}.$$
We estimate using that $\text{diam}(\mathcal{Y}) < \infty$

\[
|\chi_b(v)_i - \chi_b(v^o)_i| 
\leq |f_\eta(X_i^\top \theta) - f_{\eta^o}(X_i^\top \theta^o)|^2 
+ 2|f_\eta(X_i^\top \theta) - f_{\eta^o}(X_i^\top \theta^o)|(f_\eta(X_i^\top \theta) - f_{\eta^o}(X_i^\top \theta^o))| 
\leq 2|f_\eta(X_i^\top \theta) - f_{\eta^o}(X_i^\top \theta)|^2 + 2|f_\eta(X_i^\top \theta) - f_{\eta^o}(X_i^\top \theta^o)|^2 
+ \sqrt{2|f_\eta - \eta^o|(X_i^\top \theta)|^2 + 2|f_\eta(X_i^\top \theta) - f_{\eta^o}(X_i^\top \theta^o)|^2} 
\leq 2\|\eta - \eta^o\|^2 m^2\|\psi\|^2_\infty + 2\|\theta - \theta^o\|^2 sX^2 m^3\|\psi\|^2_\infty \|\eta^o\|^2 
+ \sqrt{2\|\eta - \eta^o\|^2 m^2\|\psi\|^2_\infty + 2\|\theta - \theta^o\|^2 sX^2 m^3\|\psi\|^2_\infty \|\eta^o\|^2} 
\leq C_1 m^2\|v - v^o\| + C_2 m^3\|v - v^o\|^2.
\]

But note that by the triangular inequality we also have $|\chi_b(v)_i - \chi_b(v^o)_i| \leq 2$. This gives

\[
\sup_{v, v^o} \frac{\|\chi_b(v) - \chi_b(v^o)\|_{L_1(P_n)}}{\|v - v^o\|_2} \leq \sup_{v, v^o} \left(\frac{2}{\|v - v^o\|_2} \wedge C_1 m^2 + C_2 m^3\|v - v^o\|_2\right) 
= C_3 m^2.
\]

We infer setting $\delta = \sqrt{p^*/n}$

\[
\frac{\sqrt{6\{1 + \log N(\delta, \mathcal{F}, L_1(P_n))\}}}{n} + \delta \leq \frac{\sqrt{6\{1 + \log N(\delta/(L(P_n) \vee 1), \mathcal{Y}, \cdot \| \cdot \|_2)\}}}{n} + \delta 
\leq \frac{\sqrt{6\{1 + \log(C_3 d\mathcal{Y}) + \log(m^2) + \log(1/\delta)p^*\}}}{n} + \delta 
\leq C_4 \sqrt{\frac{2 \log(p^*) + \log(n/p^*)p^*/2}{n}} + \sqrt{p^*/n} 
\leq C_5 \sqrt{\frac{\log(n)p^*}{n}}.
\]

We get combining (A.20) with (A.21) and (A.15) with $t = \sqrt{n}Q(2b)/2$

\[
\mathbb{P} \left(\inf_{v \in T_0(v)^c} -\mathbb{E}[\mathcal{L}(v, v^*_m)](X) < ebr^2\right) 
\leq \mathbb{P} \left(Q(2b) \leq 4e + 2/n + C_6 \sqrt{\frac{\log(n)p^*}{n}}\right) 
+ \exp \left\{-n/m^5\right\} + \exp \left\{-nQ(2b)^2/4\right\}.
\]
This gives the claim.

**Remark A.2.** By assumption \((\text{Cond}_{\mathbf{X}\theta^*})\) for any \((\theta, \eta) = v \in \mathcal{Y}\), there exist constants \(c_f', c_{d\mathbf{X}}, r_f' > 0\) and a value \((x_0, y_0) \in \{x^2 + y^2 \leq s\mathbf{X}\} \subset \mathbb{R}^2\) such that for \((x, y) \in \{(x-x_0)^2 + (y-y_0)^2 \leq r_f^2\}\) we have \(|f_{\eta'}'(x)| > c_f'\) and \(g_{\theta^*, \eta'}(x, y) \geq c_{d\mathbf{X}}\). This is where the constants in the formulation of the following lemma come from. Further \(\lambda_e \geq \mathbb{R}\) from (A.16) is strictly greater 0 because the basis functions are linearly independent and we assumed the distribution of the regressors \(\mathbf{X}\) to be absolutely continuous to the Lebesgue measure.

**Lemma A.14.** Denote \(Z_{\rho, x, y}(x_0, y_0) \overset{\text{def}}{=} \{(x, y, z) \in \mathbb{R}^2 \times \mathbb{R}^{p-2}; (x-x_0)^2 + (y-y_0)^2 \leq \rho^2\}\). There is a point \((x_0, y_0) \in \mathbb{R}^2\) such that
\[
\begin{align*}
Q(2\mathbf{b}) & \geq \frac{1}{2} \land c_{d\mathbf{X}} \lambda \left(B_{r_f', 0}\right) \\
\land \lambda \left(Z_{r_f, x, y}(0) \cap \{(x, y) \in \mathbb{R}^2 : \text{sign}(y_0)y \geq \text{sign}(y_0)r_f/2 \cap B_{s\mathbf{X}}(x_0, y_0, 0)\}\right),
\end{align*}
\]
\[
\text{for}
\]
\[
2\mathbf{b}
\]
\[
\leq \max_{c_{\theta} \in (0, 1)} \left(1 - \rho \right) \left(1 - \epsilon\right) \left(\frac{(1 - c_{\theta}) \lambda_e^2}{4\|\mathcal{D}\|^2/n \lor 1} - \frac{\sqrt{(1 - c_{\theta}) \epsilon \lambda_e L_{f'}} s\mathbf{X}}{2\|\mathcal{D}\| / \sqrt{n \lor 1} \epsilon_{CD}}\right) \\
\land \frac{c_f^2 r_f^2 c_{\theta} (1 - \rho)(1 - \epsilon)}{8p\epsilon s\mathbf{X}^2\|f\mathbf{X}\| \epsilon_{C_{\eta^*}}},
\]
\[
\text{for some small } \epsilon > 0 \text{ that decreases with the bias } \|\mathcal{D}_m(\Pi_m \mathbf{v}^* - \mathbf{v}_{m}^*)\| \text{ and for}
\]
\[
r \geq r_{Q} \overset{\text{def}}{=} \sqrt{m - \frac{2\|\mathcal{D}\| \tau(m)}{\lambda_e \sqrt{(1 - c_{\theta})(1 - \rho)(1 - \epsilon)}}},
\]
\[
\text{where for a set } A \subset \mathbb{R}^p \text{ we denote by } \lambda(A) \in \mathbb{R}_+ \text{ its Lebesgue measure and where}
\]
\[
\lambda_e \overset{\text{def}}{=} \sup \left\{ \lambda > 0 : \inf_{v \in \mathbb{R}^{p_0}, \|v\| = 1} \mathbb{P} \left(|\langle v, e(\mathbf{X}^\top \theta)\rangle| > \lambda\right) > 3/4 \right\}. \quad (A.16)
\]

**Proof.** Before we determine \(b > 0\) that allows to prove (A.14) note that
\[
\|\mathcal{D}_m(\mathbf{v} - \mathbf{v}_{m}^*)\| - \|\mathcal{D}_m(\mathbf{v}^* - \mathbf{v}_{m}^*)\| \leq \|\mathcal{D}_m(\mathbf{v} - \mathbf{v}^*)\|
\]
\[
\leq \|\mathcal{D}_m(\mathbf{v} - \mathbf{v}_{m}^*)\| + \|\mathcal{D}_m(\mathbf{v}^* - \mathbf{v}_{m}^*)\|.
\]
\[
\text{Lemma D.5 from Andresen and Spokoiny (2014) gives}
\]
\[
\|\mathcal{D}_m(\mathbf{v}^* - \mathbf{v}_{m}^*)\| \leq \sqrt{\frac{1 + \rho}{1 - \rho}} \left(\alpha(m) + \tau(m) + 2\delta(2r^*)r^*\right) \overset{\text{def}}{=} r_m^*.
\]
where due to Lemma A.2 and the definition of $r^*$ > 0 in Lemma A.3

$$r^* < \frac{8}{b} \left( C(m) \| f_X^T \theta^* \|_\infty C_{\| f^* \|} + C(m) 26\sqrt{28s_X^{p+1}} L_f X \| \psi \|_\infty C_{\| f^* \|} \right) nm^{-1/(1+2\alpha)} m,$$

$$\alpha(m) = C_1 m^{-\alpha-1/2} \sqrt{n}.$$ 

With arguments as in Lemma 3.6 we find that $r_m^* > 0$ is negligibly small for $n \in \mathbb{N}$ large enough. So we have with some small $\epsilon > 0$

$$(1 - \epsilon) \| D_m (v - v_m^*) \|^2 \leq \| D_m (v - v^*) \|^2 \leq (1 + \epsilon) \| D_m (v - v_m^*) \|^2. \quad \text{(A.17)}$$

We find for $v \in T_\epsilon(x)^c$ and with Lemma 3.4 and (A.17) that

$$\| D_{\theta \theta}^{1/2} (\varphi_\theta - \varphi_{\theta^*}) \|^2 + \| H_m (\eta - \eta^*) \|^2 \geq (1 - \rho) \| D_m (v - v^*) \|^2 \geq (1 - \rho)(1 - \epsilon)x^2.$$ 

Now we can prove (A.14). We treat two cases separately. The first case is that $\| D_{\theta \theta}^{1/2} (\varphi_\theta - \varphi_{\theta^*}) \|^2 \leq c_\theta(1 - \rho)(1 - \epsilon)x^2$. In this situation we can use the smoothness of $f_{\eta_m^*}$ and $f_{\eta^*}$ to determine $b > 0$. In the second case we use the geometric structure of

$$\left( f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta^*) \right)^2 > 0,$$

to obtain a good lower bound.

Case 1: $\| D_{\theta \theta}^{1/2} (\varphi_\theta - \varphi_{\theta^*}) \|^2 \leq c_\theta(1 - \rho)(1 - \epsilon)x^2$. In this case we simply calculate and find

$$|f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta^*)|^2 \geq |f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta)|^2 - 2|f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta)||f_{\eta^*} (X^T \theta) - f_{\eta^*} (X^T \theta^*)|
\geq |f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta)|^2 - 2|f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta)|L_f X \| \theta - \theta^* \|
\geq \left( |f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta)|^2 - L_f^2 s_X \| \theta - \theta^* \| \right)^2 - L_f^2 s_X^2 \| \theta - \theta^* \|^2.$$ 

Now

$$|f_\eta (X^T \theta) - f_{\eta^*} (X^T \theta)| \geq |f_{\eta - \eta^*} (X^T \theta)| - |f_{(0, \epsilon^*)} (X^T \theta)|.$$

We find with probability greater $3/4$

$$|f_{\eta - \eta^*} (X^T \theta)| = |\langle \eta - \eta^*, e (X^T \theta) \rangle| \geq \| \eta - \eta^* \| \| \langle v, e (X^T \theta) \rangle | \geq 1 \| H_m \| \| H_m (\eta - \eta^*) \| \| \langle v, e (X^T \theta) \rangle | \geq \frac{r \lambda_e}{\| D \| \sqrt{(1 - c\theta)(1 - \rho)(1 - \epsilon)},}$$
where

\[
\lambda_e \overset{\text{def}}{=} \sup \left\{ \lambda > 0 : \inf_{v \in \mathbb{R}^m, \|v\| = 1} \mathbb{P} \left( \|v, e(X^\top \theta)\| > \lambda \right) > 3/4 \right\},
\]

which is larger than 0 because the basis functions are linearly independent and we assumed the distribution of the regressors \(X\) to be absolutely continuous to the Lebesgue measure. Remember that by Lemma A.2

\[
\|H^{1/2}_m \mathbf{x}^*\|^2 < \left( C(m)\|f_X^\top \theta^*\|_\infty \|f^*\| + C(m)26\sqrt{28s_{X}^\top L_fX \|\psi\|_\infty \|f^*\|} \right) nm^{-2\alpha}
\]

\[\overset{\text{def}}{=} c_{\mathbf{x}}^2 m.\]

We use the Markov inequality to obtain

\[
\mathbb{P} \left( |f_{0,c^*}(X^\top \theta)|^2 \geq 4\tau(m)^2 \frac{m}{n} \right) \leq \frac{\|H^{1/2}_m \mathbf{x}^*\|^2}{4c_{\mathbf{x}}^2 m} \leq 1/4.
\]

This gives that with probability greater than \(1/2 = 3/4 - 1/4\)

\[
|f_\eta(X^\top \theta) - f_\eta^*(X^\top \theta)| \geq \frac{\tau \lambda_e \|D\| \sqrt{(1 - c_\theta)(1 - \rho)(1 - e)}}{2\|\|D\| \vee \sqrt{n}} - c_\mathbf{x} \sqrt{\frac{m}{n}}
\]

\[\geq \frac{\sqrt{(1 - c_\theta)(1 - \rho)(1 - e)}\lambda_e}{2\|\|D\| \vee \sqrt{n}} \overset{\text{def}}{=} \beta_\eta(c_\theta)r,
\]

for

\[r \geq \sqrt{\frac{2c_\mathbf{x}}{\lambda_e \sqrt{(1 - c_\theta)(1 - \rho)(1 - e)}}}.
\]

We still have to account for the summand \(L_{f^*} s_X \|\theta - \theta^*\|\) via

\[
L_{f^*} s_X \|\theta - \theta^*\| \leq \frac{L_{f^*} s_X \sqrt{c_\theta(1 - \rho)(1 - e)}}{c_D} \overset{\text{def}}{=} \beta_{\theta}(\sqrt{n}c_\theta)r.
\]

So we obtain in case 1 that \(Q(2b) \geq 1/2\) for

\[
2b/n \overset{\text{def}}{=} \beta_\eta(c_\theta)^2 - \beta_{\theta}(c_\theta)\beta_\eta(c_\theta)
\]

\[= (1 - \rho)(1 - e) \left( (1 - c_\theta)^2 (\|D\| \vee 2) / n \right) - \lambda_e L_{f^*} s_X \sqrt{c_\theta(1 - \rho)(1 - e)} \overset{\text{def}}{=} \beta_{\theta}(\sqrt{n}c_\theta)r.
\]

Case 2: \(c_\theta(1 - \rho)(1 - e)r^2 \leq \|\|D\|_{\theta\theta}^{1/2} (\varphi_\theta - \varphi_{\theta^*})\|\|^2 \leq \sqrt{2}\lambda_{\max} D^2\).

Take some \(f : \mathbb{R} \to \mathbb{R}\) with \(f' > c\) and some \((\alpha, \beta) \in \mathbb{R}^2\) with \(\alpha^2 + \beta^2 = 1\). Further take any \(g : \mathbb{R} \to \mathbb{R}\). We denote \(Z_{\rho,x,y}(x_0, y_0) \overset{\text{def}}{=} \{(x, y, z) \in \mathbb{R}^2 \times \mathbb{R}^{p-2}; (x - x_0)^2 + \ldots\}\).
\((y - y_0)^2 \leq \rho^2\). We are interested in determining
\[
V(\tau) \overset{\text{def}}{=} \inf_{f \in C^1(\mathbb{R}), f' > c, \ g: \mathbb{R} \to \mathbb{R}} \lambda \left( \{ |f(\alpha x + \beta y) - g(x)| > \tau \} \cap Z_{\rho, x, y}(0) \cap B_{sX}(x_0, y_0, 0) \right),
\]
where for a set \(A \subset \mathbb{R}^p\) we denote by \(\lambda(A) \in \mathbb{R}_+^\ast\) its Lebesgue measure. For this observe
\[
f(\alpha x + \beta y) - g(x) \begin{cases} 
\geq c \beta y + f(\alpha x) - g(x), & \beta > 0, \\
\leq c \beta y + f(\alpha x) - g(x), & \beta \leq 0
\end{cases}
\]
Consequently for fixed \(x \in [-\rho, \rho]\) we have \(|f(\alpha x + \beta y) - g(x)| > \rho \beta c/2\) on the set
\[
\{y \in [-\sqrt{\rho^2 - x^2}, \sqrt{\rho^2 - x^2}] : |c \beta y + f(\alpha x) - g(x)| > \rho \beta c/2\},
\]
which always is of a length greater \(\lambda([-\sqrt{\rho^2 - x^2}, \sqrt{\rho^2 - x^2}] \setminus [-\rho/2, \rho/2])\). Addressing the way a centered cylinder intersects with a shifted ball this gives that
\[
V(\rho \beta c/2) \geq \lambda \left( Z_{\rho, x, y}(0) \cap \{(x, y) \in \mathbb{R}^2 : -\text{sign}(y_0) y \geq -\text{sign}(y_0) \rho/2\} \cap B_{sX}(x_0, y_0, 0) \right).
\]
Now we can prove the claim. For any \((\theta, \eta) = \nu \in \mathcal{Y}\), with \(\|\theta\| = 1\), we can represent \(\theta^* = \alpha \theta + \beta \theta^0\) with some \(\theta^0 \in \theta^\perp\) with \(\|\theta^0\| = 1\) and \(\alpha^2 + \beta^2 = 1\). By assumption \((\text{Cond}_X \theta^*)\) for any \((\theta, \eta) = \nu \in \mathcal{Y}\), there exist constants \(c_f, c_d, r_f > 0\) and a value \((x_0, y_0) \in \{x_0^2 + y_0^2 \leq s_X\} \subset \mathbb{R}^2\) such that for \((x, y) \in \{(x - x_0)^2 + (y - y_0)^2 \leq r_f^2\}\) we have \(|f'_{\eta}(x)| > c_f\) and \(d_X \geq c_d\). We can estimate writing \(Z_{r_f, x, y}(x_0, y_0) \overset{\text{def}}{=} \{(x, y, z) \in \mathbb{R}^2 \times \mathbb{R}^{d_f} ; (x - x_0)^2 + (y - y_0)^2 \leq r_f^2\}\)
\[
\mathbb{P}\left\{(f_{\eta^*}(X^\top \theta^*) - f_{\eta}(X^\top \theta))^2 \geq c_f^2 r_f^2 \beta^2 / 4\right\}
\geq \inf_{f \in C^1(\mathbb{R}), f' > 0, \ g: \mathbb{R} \to \mathbb{R}} \mathbb{P}(\{ |f(\alpha x + \beta y) - g(x)| \geq c_f r_f \beta / 2 \} \cap \{ \theta : \{ \theta \in \mathbb{R}^d \} \cap \{ \theta \in \mathbb{R}^d \} \}
\cap \{ X \in B_{sX}(0) \})
\geq c_d \inf_{f \in C^1(\mathbb{R}), f' > 0, \ g: \mathbb{R} \to \mathbb{R}} \lambda \left( \{ |f(\alpha x + \beta y) - g(x)| \geq c_f r_f \beta / 2 \} \cap \{ \theta : \{ \theta \in \mathbb{R}^d \} \cap \{ \theta \in \mathbb{R}^d \} \}
\cap \{ X \in B_{sX}(0) \}
\geq c_{dX} \lambda \left( Z_{r_f, x, y}(0) \cap \{ \theta : \{ \theta \in \mathbb{R}^d \} \cap \{ \theta \in \mathbb{R}^d \} \}
\cap \{ X \in B_{sX}(0) \}
\geq c_{dX} \lambda \left( Z_{r_f, x, y}(0) \cap \{ \theta : \{ \theta \in \mathbb{R}^d \} \cap \{ \theta \in \mathbb{R}^d \} \}
\cap \{ X \in B_{sX}(0) \}\right).
\]
We need to express \(\beta > 0\) in terms of \(r > 0\). Since \(\alpha = \sqrt{1 - \beta^2}\)
\[
\left(1 - \sqrt{2 - \beta^2}\right)^2 + \beta^2 = (1 - \alpha)^2 + \beta^2 = \|\theta - \theta^*\|^2.
\]
This gives that

\[ \beta^2 \geq \| \theta - \theta^* \|^2 \left( 1 - \frac{1}{4} \| \theta - \theta^* \|^2 \right). \]

Clearly on the sphere we have \( \| \theta - \theta^* \|^2 \leq 2 \) such that

\[ \beta^2 \geq \| \theta - \theta^* \|^2 / 2. \]

Further for any \( \varphi_\theta, \varphi_\theta \in W_S \) we have with (A.17) that

\[ \| \theta - \theta^* \|^2 \geq \frac{2}{p \pi^2} \| \varphi_\theta - \varphi_\theta^* \|^2 \geq \frac{2}{p \pi^2 \lambda_{\text{max}} D^2} \| D(\varphi_\theta - \varphi_\theta^*) \|^2 \geq \frac{c_\theta(1 - \rho)(1 - \epsilon)}{p \pi^2 \lambda_{\text{max}} D} r^2. \]

By Lemma A.4 we find

\[ \beta^2 \geq \frac{c_\theta(1 - \rho)(1 - \epsilon)}{8 p \pi^2 s_X^2 \| f_\theta \|^2 \| \eta^* \|} r^2 / n. \]

Combined this yields that with

\[ \frac{2b}{n} \overset{\text{def}}{=} \frac{c_p r^2}{n} \frac{c_\theta(1 - \rho)(1 - \epsilon)}{8 p \pi^2 s_X^2 \| f_\theta \|^2 \| \eta^* \|}, \]

it holds

\[ \mathbb{P} \left\{ \left( f_\eta(X^\top \theta) - f_\eta^*(X^\top \theta^*) \right)^2 \geq b r^2 / n \right\} \]

\[ \geq c_{dX} \lambda \left( B_{r^*}(0) \cap \{(x, y) \in \mathbb{R}^2 : |y| \leq r^*/2 \} \right). \]

This gives the claim. \( \square \)

**Lemma A.15.** We have for some \( C > 0 \)

\[ n \mathbb{E} \| f_\eta_m(X^\top \theta_m^*) - f_\eta^*(X^\top \theta^*) \|^2 \leq 3(2 + C) r^2 \overset{\text{def}}{=} c_m. \]

**Proof.** We find with the Taylor expansion, Lemma A.3 (which is applicable it only only needs \( \mathcal{L} \mathbf{r} \) for the full model and with center \( \mathbf{v}^* \in \mathcal{T} \), which we already proved with Lemma A.13, since \( c_m = c_{\Sigma} = 0 \) in that case) and Lemma A.2 with some
We have for a constant

\[ \xi \in \text{Conv}(\theta^*_m, \theta^*) \]

\[ \begin{align*}
& n\mathbb{E}[\|f_{\eta^*_m}(X^\top \theta^*_m) - f_{\eta^*}(X^\top \theta^*)\|^2] \\
& \quad \leq 3n \left( \mathbb{E}[\|f_{\eta^*}(X^\top \theta^*_m) - f_{\eta^*}(X^\top \theta^*)\|^2] + \mathbb{E}[\|f_{\eta^*_m} - f_{\eta^*}(X^\top \theta^*_m)\|^2] \right) \\
& \quad \leq 3 \left( \|D^{1/2}_{\theta \theta}(\xi)(\theta^*_m - \theta^*)\|^2 + \|D^{1/2}_{ff}(v^*_m)(\eta^*_m - f^*)\|^2 \right) \\
& \quad \leq 3 \left( (1 + I - D^{-1/2}n\tilde{D}(\xi)D^{-1/2})\|D(\theta^*_m - \theta^*)\|^2 \\
& \quad + (1 + I - D^{-1/2}n\tilde{D}_{ff}(v^*_m)D_{ff}^{-1/2})\|D_{ff}(\eta^*_m - f^*)\|^2 \right) \\
& \quad \leq 3 \left[ 2 + (I - D^{-1/2}n\tilde{D}(\xi)D^{-1/2}) + (I - D^{-1/2}n\tilde{D}_{ff}(v^*_m)D_{ff}^{-1/2}) \right] \\
& \quad \|D(v^*_m - v^*)\|^2 \\
& \quad \leq 3(2 + c)x^2.
\end{align*} \]

\[ \Box \]

**Lemma A.16.** We have for a constant $c_1 > 0$ that only depends on $\|\psi\|_\infty, \|\psi'\|_\infty, sX^2$ that

\[ \begin{align*}
& \mathbb{P} \left( n \left| (P_n - \mathbb{P})f_{\eta^*_m}(X^\top \theta^*_m) - f_{\eta^*}(X^\top \theta^*) \right|^2 \right) \geq c_1 x^* m^{1/2} \right) \leq \exp \left\{ -n/m^5 \right\}.
\end{align*} \]

**Proof.** We want to use the finite difference inequality to show that $L(P_n) - \mathbb{E}L(P_n)$ is small with high probability. As above define

\[ f : \bigotimes_{i=1}^n \mathbb{R}^p \to \mathbb{R}, \quad f(X_1, \ldots, X_n) \overset{\text{def}}{=} P_n|f_{\eta^*_m}(X^\top \theta^*_m) - f_{\eta^*}(X^\top \theta^*)|^2, \]

and note that for any $i = 1, \ldots, n$ and any alternative realization $X'_i \in \mathbb{R}$

\[ n|f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)| \]

\[ \leq |f_{\eta^*_m}(X^\top_i \theta^*_m) - f_{\eta^*}(X^\top_i \theta^*)|^2 + |f_{\eta^*_m}(X^\top_i \theta^*_m) - f_{\eta^*}(X^\top' \theta^*)|^2. \]
Further with A.3

\[
|f_{\eta_m^*}(X^T\theta^*) - f_{\eta^*}(X^T\theta^*)|^2 \\
\leq 2|f_{\eta_m^*}(X^T\theta)|^2 + 2|f_{\eta^*}(X^T\theta^*) - f_{\eta^*}(X^T\theta^*)|^2 \\
+ \sqrt{2|f_{\eta_m^*}(X^T\theta)|^2 + 2|f_{\eta^*}(X^T\theta^*) - f_{\eta^*}(X^T\theta^*)|^2} \\
+ \sqrt{2|\eta^* - \eta_m^*|^2 + 2|\theta^* - \theta_m^*|^2 s^2 m^2 \|\psi\|^2 \|\eta^*\|^2} \\
+ \sqrt{m\|\psi\|_\infty^2 (\|\eta^*\|^2)} \\
\leq C_1 \frac{m^3}{n c_D} \|D(v^* - v_m^*)\|
\]

for a constant $C_1 > 0$. Note that with Lemma A.2 of Andresen (2014) combined with Lemma A.2 we find $\|D(v^* - v_m^*)\| \to 0$. Assume that $n, m \in \mathbb{N}$ are large enough to ensure that $\|D(v^* - v_m^*)\| \leq 1$. This gives with the bounded difference inequality (Azuma Hoeffding inequality) that

\[
P\left(\left|\left(P_n - P\right)|f_{\eta_m^*}(X^T\theta^*) - f_{\eta^*}(X^T\theta^*)|^2 \right| \geq tC_1 m^3 / n^{3/2}\right) \leq \exp \{-t^2\},
\]

From this we infer with $t = \sqrt{n}/m^{5/2} \to \infty$

\[
P\left(n \left|\left(P_n - P\right)|f_{\eta_m^*}(X^T\theta^*) - f_{\eta^*}(X^T\theta^*)|^2 \right| \geq C_1 m^1 / 2\right) \leq \exp \{-n / m^5\}.
\]

To summarize we get with Lemma A.14

\[
Q(2b) \geq \frac{1}{2} \wedge c_{dX} \lambda \left( B_{r_f}(0) \cap \{ (x, y) \in \mathbb{R}^2 : |y| \leq r_f \beta c_f / 2 \} \right),
\]

if $b > 0$ with

\[
2b \leq \max_{c_{\theta} \in (0, 1)} \left( (1 - \rho)(1 - \epsilon) \left( 1 - c_{\theta} \right) \frac{\lambda^2}{4\|D\|^2} - \sqrt{\left( 1 - c_{\theta} \right) c_{\theta} \lambda e L_f s X} \frac{2\sqrt{n\|D\|c_D}}{2\sqrt{n\|D\|c_D}} \right) \\
\wedge \frac{c_{r_f}^2 r_f^2 c_{\theta} (1 - \rho)(1 - \epsilon)}{8 p^2 s X^2 \|f_X\|^2 \|\eta\|^2 C_{\|\eta\|^2}}
\]

for some small $\epsilon > 0$ that decreases with the bias $\|D_m(P_m v^* - v_m^*)\|$ and for

\[
r \geq rQ \overset{\text{def}}{=} \sqrt{m \frac{2\|D\|\tau(m)}{\lambda e (1 - c_{\theta})(1 - \rho)(1 - \epsilon)}}.
\]
Finally Lemma A.15 and Lemma A.16 give that for $n \in \mathbb{N}$ large enough with probability $1 - \exp\{-n/m^5\}$

$$C_m = 3(2 + C)r^2, \quad C_\Sigma = C_1 r^* m^{1/2}. $$

Plugging these into Lemma A.13 gives the claim since $r^* \leq C \sqrt{m}$.  

\[\square\]

A.3.4 Proof of the second regularity condition

This is a direct consequence of the following lemma:

**Lemma A.17.** For a positive definite symmetric matrix

$$
\mathcal{D} = \begin{pmatrix}
D^2 & A \\
A^\top & H^2
\end{pmatrix},
$$

with $c_D \|v\|^2 \leq v^\top \mathcal{D} v$ for some $c_D > 0$ we have that

$$\|D^{-1}AH^{-2}A^\top D^{-1}\| =: \rho^2 \leq 1 - \frac{c_D}{\lambda_{\max} D \wedge \lambda_{\max} H}.$$

**Proof.** For any $v = (\theta, \eta) \in \mathbb{R}^{p+m}$ we have

$$v^\top \mathcal{D} v = (\theta^\top, \eta^\top) \begin{pmatrix}
D^2 & A \\
A^\top & H^2
\end{pmatrix} \begin{pmatrix}
\theta \\
\eta
\end{pmatrix} = (\theta^\top D^\top, \eta^\top H^\top) \begin{pmatrix}
I_p & D^{-1}AH^{-1} \\
H^{-1}A^\top D^{-1} & I_m
\end{pmatrix} \begin{pmatrix}
D\theta \\
H\eta
\end{pmatrix} = \|D\theta\|^2 + \|H\eta\|^2 + 2\langle H\eta, H^{-1}A^\top D^{-1}\theta \rangle.$$

This gets minimal with $H\eta = -H^{-1}A^\top D^{-1}D\theta$. In that case

$$v^\top \mathcal{D} v = \|D\theta\|^2 - \|H^{-1}A^\top D^{-1}D\theta\|^2 = (D\theta)^\top (I_p - D^{-1}AH^{-2}A^\top D^{-1})D\theta,$$

which gets minimal if

$$D^{-1}AH^{-2}A^\top D^{-1}D\theta = \|D^{-1}AH^{-2}A^\top D^{-1}\|D\theta = \rho^2 D\theta,$$

i.e. if $D\theta \in \mathbb{R}^p$ is a maximal eigenvalue of $D^{-1}AH^{-2}A^\top D^{-1} \in \mathbb{R}^{p \times p}$. With the assumption $c_D \|v\|^2 \leq v^\top \mathcal{D} v$ this gives

$$c_D \|v\|^2 \leq v^\top \mathcal{D} v = (1 - \rho^2)\|D\theta\|^2, \quad \|v\|^2 = \|\theta\|^2 + \|H^{-2}A^\top \theta\|^2,$$

such that

$$\rho^2 \leq 1 - c_D \frac{\|\theta\|^2}{\|D\theta\|^2} \leq 1 - \frac{c_D}{\lambda_{\max} D}.$$
With analogous arguments we can obtain

\[ \rho^2 \leq 1 - c_D \frac{\| \eta \|^2}{|H \eta|^2} \leq 1 - \frac{c_D}{\lambda_{\text{max}} H}, \]

which completes the proof.

\[ \square \]

### A.4 Proof of Lemma 3.5

**Proof.** As \( b(r) = b = \frac{c(L)}{2} \) is a constant in our case we can use Theorem C.1 of Andresen and Spokoiny (2014). A thorough glance at the details of the paper Spokoiny (2012) reveals that the size of \( r_0 > 0 \) is determined by \( \nu_{r,m}^2 \). Let \( r_0 > 0 \) such that the second condition of the mentioned Theorem C.1 (adapted to the case \( \nu_{r,m}^2 \neq \nu_{1,m} \)) is satisfied for all \( r \geq r_0 \):

\[ 6 \nu_{r,m} \sqrt{x + Q} \leq r \frac{c(Lr)}{2}, \]

which means that \( r_0 = 12 \frac{\nu_{r,m}}{c(Lr)} \sqrt{x + Q} \). Because \( Q = 4p^* \) in our setting and with Lemma 3.3 we find

\[ r_0 = 12 \frac{\nu_{r,m}}{c(Lr)} \left( c_M^2 + c_M^2 (x^o) m^3 / \sqrt{n} \right)^{1/2} \sqrt{x + 4p^*}. \]

Now we treat the first condition of Theorem C.1 of Andresen and Spokoiny (2014) i.e. we want to satisfy

\[ 1 + \sqrt{x + Q} \leq 3 \nu_{r,m} g(r) / b. \]

Remember that

\[ g(r) = c_D \tilde{g} C (x^o)^{-1} m^{-3/2} \sqrt{n}. \]

which gives that the second condition is always satisfied if

\[ 1 \leq \frac{6 c_D \tilde{g} \nu_{r,m}^2 \left( c_M^2 + c_M^2 (x^o) m^3 / \sqrt{n} \right) \sqrt{n}}{m^{3/2} c(Lr) (1 + \sqrt{x + 4m})}. \]

such that the second condition is always satisfied for \( n \in \mathbb{N} \) large enough. \[ \square \]

### A.5 Proof of Lemma 3.6

**Proof.** First note that by the definition of \( j(x, Q) \) we have for moderate \( x > 0 \)

\[ \diamondsuit(r) \leq \left( \delta(r) + 6 \sqrt{8} \nu_{1} \omega \sqrt{p^* + x} \right) r, \]
where with Lemma 3.3
\[ \delta(r) = \frac{C(L_0) \left( m^{3/2} \sqrt{\frac{m^3}{\sqrt{n}}} \right) [r + r^*]}{c_D \sqrt{n}}, \]
\[ \omega \overset{\text{def}}{=} \frac{2}{\sqrt{nc_D}}, \]
\[ \nu^2_{1,m} = \bar{\nu}^2 r m^3 C(\mathcal{E}_D) \left( \frac{1}{c_D^2} + \frac{\sqrt{5}(x + \log(2m))^{1/2}}{\sqrt{n}} \right)^{1/2}. \]

Putting this together we find
\[ \diamondsuit(r) = \frac{C(L_0) \left( m^{3/2} \sqrt{\frac{m^3}{\sqrt{n}}} \right) [r^2 + r^* r]}{c_D \sqrt{n}} \]
\[ + 12 \sqrt{8} \frac{\bar{\nu} r m^3 C(\mathcal{E}_D)}{c_D \sqrt{n}} \left( \frac{1}{c_D^2} + \frac{\sqrt{5}(x + \log(2m))^{1/2}}{\sqrt{n}} \right)^{1/2} \sqrt{x + p^* r} \]

Using Lemma 3.5 we find that
\[ \frac{r_0 m^{3/2}}{\sqrt{n}} \leq 12 \frac{\bar{\nu} r}{c(L_0)} \left( C^2_{\mathcal{D}r_0} + C^2_{\mathcal{E}x} m^3/\sqrt{n} \right)^{1/2} \sqrt{x + 4 p^* m^{3/2}} \sqrt{n} \]
\[ \leq 12 \frac{\bar{\nu} r}{c(L_0)} \left( C_{\mathcal{D}r_0} (r^0)^2 + C^2_{\mathcal{E}x} \right)^{1/2} \sqrt{x + 4 \left( \frac{p^* m^{3/2}}{n^{3/4}} \right)}, \]

which means that if \( n_0 \in \mathbb{N} \) is large enough to ensure that for all \( n \geq n_0 \)
\[ \frac{p^* m^{3/2}}{n^{3/4}} \leq \left( 12 \frac{\bar{\nu} r}{c(L_0)} \left( C_{\mathcal{D}r_0} (r^0)^2 + C^2_{\mathcal{E}x} \right)^{1/2} \sqrt{x + 4} \right)^{-1}, \]
\[ \frac{\sqrt{5} m^{3/2} \left( x + \log(2m) \right)^{1/2}}{\sqrt{n}} \leq C_{\mathcal{E}_x}, \]

we can estimate
\[ \diamondsuit(r_0) \leq C(L_0) \frac{m^{3/2} [r_0^2 + r^* r_0]}{c_D \sqrt{n}} \]
\[ + 12 \sqrt{8} \bar{\nu} r C(\mathcal{E}_D) \left( \frac{1}{c_D^2} + \frac{\sqrt{5}(x + \log(2m))^{1/2}}{\sqrt{n}} \right)^{1/2} \frac{m^{3/2} \sqrt{x + p^* r_0}}{c_D \sqrt{n}}. \]
Remember that 
\[
r^* = 16 \left( C(m) \| f_X \theta^* \|_\infty C\| f^* \| + C(m) 26 \sqrt{28 s_X^{p+1}} L_{f_X} \| \psi \|_\infty C^2\| f^* \| \right)^{1/2} \sqrt{\frac{nm^{-1+(1+2\alpha)/2}}{\sqrt{m}}}. 
\]

Consequently as soon as \( n_1 \in \mathbb{N} \) is large enough to ensure that for all \( n \geq n_1 \vee n_0 \)
\[
m^{-1+(1+2\alpha)/2} \leq 16 \left( C(m) \| f_X \theta^* \|_\infty C\| f^* \| + C(m) 26 \sqrt{28 s_X^{p+1}} L_{f_X} \| \psi \|_\infty C^2\| f^* \| \right)^{1/2},
\]
we get for \( n \geq n_1 \vee n_0 \) that \( [r^2 + r^* r_0] \leq 2r_0^2 \). We find for all \( n \geq n_1 \vee n_0 \)
\[
\Diamond (r_0) \leq C\Diamond \frac{(x + p^*)^{5/2}}{\sqrt{n}},
\]
with some appropriate constant \( C\Diamond > 0 \). So for \( n \geq n_1 \vee n_0 \) large enough to ensure \( C\Diamond (x + p^*)^2/\sqrt{n} \leq 1 \) we get
\[
\Diamond (r_0) \leq \sqrt{p^* + x}.
\]

This gives the first claim. To prove the second claim note that since \( V^2 = \sigma^2 \mathcal{D} \) we can bound for moderate \( x \leq x_c \) (see Appendix C of Andresen and Spokoiny (2014) for more details)
\[
\tilde{\zeta}(x, \mathcal{D}^{-1} V^2 \mathcal{D}^{-1}) \leq \sigma \sqrt{6} \sqrt{p^* + x}.
\]

Plugging this into the definition of \( r_1 > 0 \) in combination with eq: bound for excgr rups in proof of bound for excgr this gives the claim with an appropriate constant \( C\Diamond > 0 \).

**A.6 Proof of Proposition 2.1**

*Proof.* Due to Lemma 3.3 and A.17 we may apply apply Theorem 2.1 of Andresen and Spokoiny (2014). Since \( \tilde{V}^2 = \sigma^2 \tilde{\mathcal{D}} \) we can bound for moderate \( x \leq x_c \) (see Appendix C of Andresen and Spokoiny (2014) for more details)
\[
\tilde{\zeta}(x, \tilde{\mathcal{D}}^{-1} \tilde{V}^2 \tilde{\mathcal{D}}^{-1}) \leq \sigma \sqrt{6} \sqrt{p^* + x}.
\]

So we get the claim again noting that by Lemma 3.6
\[
\Diamond (r_0, x) \leq C\Diamond (p^* + x)^{5/2}/\sqrt{n},
\]
and via adapting the size of \( C\Diamond > 0 \). \(\square\)
A.7 Proof of Lemma 3.9

Proof. It suffices to show that

$$\text{Cov}(\nabla_{\theta} (\ell_i(v^*_m) - \ell_i(v^*))) \to 0, \quad \text{Cov}(\nabla_{(\eta_1, \ldots, \eta_m)} (\ell_i(v^*_m) - \ell_i(v^*))) \to 0.$$ 

We calculate

$$\|\text{Cov}(\nabla_{\theta} (\ell_i(v^*_m) - \ell_i(v^*)))\| \leq E\| \left( f'_{\eta_m^*}(X_i^\top \theta^*_m) - f'_{\eta^*}(X_i^\top \theta^*) \right) \nabla \Phi(\theta)^\top X_i \|^2$$

$$\leq s^2 \sum_{k=0}^{\infty} \|f'_{\eta_m^*}(X_i^\top \theta^*_m) - f'_{\eta^*}(X_i^\top \theta^*)\|^2$$

$$\leq 4s^2 \sum_{k=0}^{\infty} \|e'_{k}\|_\infty (\eta^*_{m_k} - \eta^*_{k})^2 + 4s^4 \sum_{k=0}^{\infty} \|e'_{k}\|_\infty (\eta^*_{m_k} - \eta^*_{k}) \| \theta_m^* - \theta^* \|^2.$$

We estimate separately

$$\sum_{k=0}^{\infty} \|e'_{k}\|_\infty (\eta^*_{m_k} - \eta^*_{k}) \leq \|\psi'\|_\infty \left( \sum_{k=0}^{m-1} k^{3/2} (\eta^*_{m_k} - \eta^*_{k}) + \sum_{k=m}^{\infty} k^{3/2} \eta^*_{k} \right)$$

$$\leq \|\psi'\|_\infty \left( m^2 \|\eta^*_{m} - \eta^*\| + \left( \sum_{k=m}^{\infty} k^{-2\alpha-4} \right)^{1/2} \right) \left( \sum_{k=m}^{\infty} 2^\alpha \eta^2_{k} \right)^{1/2}$$

$$\leq \|\psi'\|_\infty \left( m^2 \frac{1}{\sqrt{nC_D}} \|D_m(v^* - v^*_m)\| + \sqrt{(2\alpha - 3)/(2\alpha - 4)} \left( \sum_{k=m}^{\infty} 2^\alpha \eta^2_{k} \right)^{1/2} \right)$$

The last term tends to 0 because of Lemma A.2 of Andesen (2014) and because $\sum_k 2^\alpha \eta^2_k < \infty$. Further we get with similar steps

$$\left( \sum_{k=0}^{m-1} \|e'_{k}\|_\infty \eta^*_{m_k} \right) \| \theta^*_m - \theta^* \| \leq \|\psi''\|_\infty \| \theta^*_m - \theta^* \| \left( \sum_{k=0}^{m-1} k^{5/2} \eta^*_{m_k} \right)$$

$$\leq \|\psi''\|_\infty \| \theta^*_m - \theta^* \| \left( \sum_{k=0}^{m-1} k^{2\alpha-5} \right)^{1/2} \left( \sum_{k=0}^{m-1} k^{2\alpha} \eta^2_{k} \right)^{1/2}$$

$$+ \frac{1}{\sqrt{nC_D}} \left( \sum_{k=0}^{m-1} k^{5} \right)^2 \|D_m(v^* - v^*_m)\|$$

$$\leq \|\psi''\|_\infty \| \theta^*_m - \theta^* \| \left( mC_{\eta^*} + \frac{1}{\sqrt{nC_D}} m^3 \|D_m(v^* - v^*_m)\| \right)$$

$$\leq \|D_m(v^* - v^*_m)\| \frac{1}{\sqrt{nC_D}} mC_{\eta^*} \|\psi''\|_\infty + \frac{1}{nC_D} m^3 \|D_m(v^* - v^*_m)\|^2 \|\psi''\|_\infty.$$
Again the last term tends to 0. Similarly we calculate
\[
\text{Cov}(\nabla_{(\eta_1, \ldots, \eta_m)} (\ell_i(v^*_m) - \ell_i(v^*))) \leq \mathbb{E}\|e(X_i^\top \theta^*_m) - e(X_i^\top \theta^*)\|^2
\]
\[
\leq s^2_X \|\psi\|_\infty^2 \|\theta^*_m - \theta^*\|^2 \left( \sum_{k=0}^{m-1} k^{3/2} \right)^2
\]
\[
\leq s^2_X \|\psi\|_\infty \frac{1}{\gamma C_D} m^5 \|D(v^* - v^*_m)\|^2,
\]
which again is a zero sequence. This gives the claim. \(\square\)

A.8 Auxiliary Theorems

Theorem A.18. Let a function \(f : \mathcal{X}^n \to \mathbb{R}\) satisfy for any \(X_1, \ldots, X_n, X'_i \in \mathcal{X}\)
\[
|f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)| \leq c_i.
\]
Then for any vector of independent random variables \(X \in \mathcal{X}^n\)
\[
\mathbb{P}(f(X) - \mathbb{E}f(X) \geq t) \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}, \quad (A.18)
\]
\[
\mathbb{P}(f(X) - \mathbb{E}f(X) \leq -t) \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}.
\]

Theorem A.19. Let for a sequence of independent \(X_i \in \mathcal{X}\) for some space \(\mathcal{X}\)
\[
F(v) = \sum_{i=1}^{n} f_i(v, X_i) + g, \quad v \in \mathcal{T} \subset \mathbb{R}^p
\]
and assume that with \(r > r_Q > 0, \mathcal{T}_0(r) \subset \mathcal{T}\) and \(\chi_b : [0, 2b] \to \mathbb{R}\) defined in (A.19)
\[
\mathbb{E} \left[ \sup_{v \in \mathcal{T}_0(r)^c} (P_n - \mathbb{P}) \chi_b(v) \right] \leq C_X, \quad \mathbb{P}(g > C_g) \leq \tau_g,
\]
\[
Q(b) \overset{\text{def}}{=} \inf_{v \in \mathcal{T}_0(r)^c} \mathbb{P}(f_i(v, X_i) \geq br^2/n) > 0.
\]
Choose
\[
0 < \lambda \leq (Q(2b) - 2/n + 2C_X) / 4.
\]
Then for \(r^2 \geq C_g / (\lambda b) \lor r_Q^2\)
\[
\mathbb{P} \left( \inf_{v \in \mathcal{T}_0(r)^c} F(v) \leq \lambda br^2 \right) \leq \exp \{-nQ(2b)^2/4\} + \tau_g
\]
Remark A.3. In the proof we follow closely the proof of Theorem 4.3 of Mendelson (2014). Our set $\mathcal{Y}_c(r) \subset \mathbb{R}^p$ is neither star shaped, nor convex but we can still use the same arguments.

Proof. Denote 

$$Z_i(\nu, b) = 1_{\{f_i(\nu) \geq br^2\}}.$$ 

With (A.12) and (A.13) and $r^2 \geq (C_m + C_m)/(b\epsilon) \lor r_Q$ we get

$$
\begin{align*}
\mathbb{P} \left( \inf_{\nu \in \mathcal{Y}_c(r)} F(\nu) < \epsilon b r^2 \right) \\
\leq \mathbb{P} \left( \inf_{\nu \in \mathcal{Y}_c(r)} \sum_{i=1}^n f_i(\nu) < \epsilon b r^2 + C_m + C_m \right) + \tau_g \\
\leq \mathbb{P} \left( \inf_{\nu \in \mathcal{Y}_c(r)} |\{i = 1, \ldots, n : Z_i(\nu, b) = 1\}| \leq \lceil 2\epsilon n \rceil \right) + \tau_g.
\end{align*}
$$

Define the following auxiliary function

$$\chi_u(t) = \begin{cases} 
0 & t \leq u; \\
t/u - 1 & t \in [u, 2u]; \\
1 & t \geq 2u;
\end{cases} \quad \chi_b(\nu)^i \overset{\text{def}}{=} \chi_b(f_i(\nu)),$$

and note that $Z_i(\nu, b) \geq \chi_b(\nu)^i$ while $Z_i(\nu, 2b) \leq \chi_b(\nu)^i$. This gives for $r > r_Q > 0$ using (A.14)

$$
\begin{align*}
\inf_{\nu \in \mathcal{Y}_c(r)} |\{i = 1, \ldots, n : Z_i(\nu) = 1\}| &= \inf_{\nu \in \mathcal{Y}_c(r)} n P_n Z(\nu) \\
&\geq n Q(2b) - n \sup_{\nu \in \mathcal{Y}_c(r)} \{P_n Z(\nu, b) - \mathbb{P} Z(\nu, 2b)\} \\
&\geq n Q(2b) - n \sup_{\nu \in \mathcal{Y}_c(r)} (P_n - \mathbb{P}) \chi_b(\nu),
\end{align*}
$$

where we used the shorthand notation $P_n h = \frac{1}{n} \sum_{i=1}^n h(X_i)$ for a function $h : \mathbb{R}^p \to \mathbb{R}$. So we infer

$$
\begin{align*}
\mathbb{P} \left( \inf_{\nu \in \mathcal{Y}_c(r)} -\mathbb{E}[\mathcal{L}(\nu, \nu_m^*)(X)] < \epsilon b r^2 \right) \\
\leq \mathbb{P} \left( Q(2b) - \sup_{\nu \in \mathcal{Y}_c(r)} (P_n - \mathbb{P}) \chi_b(\nu) \leq 2\epsilon + 1/n \right) + \exp\left\{-n/m^5\right\}.
\end{align*}
$$

Define

$$f : \mathcal{X}^n \to \mathbb{R}, \quad f(X_1, \ldots, X_n) \overset{\text{def}}{=} \sup_{\nu \in \mathcal{Y}_c(r)} (P_n - \mathbb{P}) \chi_b(\nu),$$
Finite sample analysis of profile M-estimation in the Single Index model

and note that for any $i = 1, \ldots, n$ and any alternative realization $X'_i$

$$|f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)| \leq 2/n.$$  

This gives with (one sided) the bounded difference inequality (Azuma Hoeffding inequality) that

$$I_P \left( \sup_{\nu \in T_\nu(x)^c} (P_n - IP) \chi_b(\nu) - IE \left[ \sup_{\nu \in T_\nu(x)^c} (P_n - IP) \chi_b(\nu) \right] \right) \geq \frac{t}{\sqrt{n}} \right) \leq e^{-t^2}. \quad (A.21)$$

By assumption

$$IE \left[ \sup_{\nu \in T_\nu(x)^c} (P_n - IP) \chi_b(\nu) \right] \leq C_\chi.$$

We get combining (A.20) with (A.21) and (A.15) with $t = \sqrt{nQ(2b)/2}$

$$I_P \left( \inf_{\nu \in T_\nu(x)^c} -IE[\mathcal{L}(\nu, \nu^*_m)(X)] < \epsilon b \sigma^2 \right)$$

$$\leq I_P (Q(2b) \leq 4\epsilon + 2/n + 2C_\chi)$$

$$+ \exp \left\{ -n/m^5 \right\} + \exp \left\{ -nQ(2b)^2/4 \right\}.$$

This gives the claim.  

B Conditions

We adopt the conditions from Section 3 of Spokoiny (2012) with some minor changes. First we present the parametric conditions that apply to parametric models with finite dimensional parameter. Then explain two new conditions that arise in the infinite dimensional setting.

For some finite dimension $p^* \in \mathbb{N}$ the parametric conditions involve two positive definite matrices, the information matrix $\mathcal{D}_0^2$ and the covariance $\mathcal{V}_0^2$ and a central point $\nu^* \in \mathbb{R}^{p^*}$ that have to be specified before the conditions can be checked.

Remark B.1. For Theorem 3.3 the matrices equal with some $\nu^* \in \mathbb{R}^{p^*} \times S$ we impose

$$\mathcal{V}_0^2 = \Pi_\nu \text{Cov} \left( \nabla \mathcal{L}(\nu^*) \right) \Pi_\nu^\top, \quad \mathcal{D}^2 = -\Pi_\nu \nabla^2 IE \mathcal{L}(\nu^*) \Pi_\nu^\top,$$

where it is important to note that in that case $\nu^0 = \nu^*_m$ where in general we have to expect that $\nu^*_m \neq \Pi_\nu \nu^*$. 

The matrices $D^2_0$ and $V^2_0$ have to satisfy certain regularity conditions. We begin by representing the information and the covariance matrices in block form:

$$
D^2_0 = \begin{pmatrix} D^2_0 & A_0 \\ A_0^\top & H^2_0 \end{pmatrix}, \quad V^2_0 = \begin{pmatrix} V^2_0 & B_0 \\ B_0^\top & Q^2_0 \end{pmatrix}.
$$

Here we restate identifiability conditions:

- **(I)** There is a constant $a > 0$ such that
  
  $$a^2 D^2_0 \geq (V_0)^2, \quad a^2 H^2_0 \geq (Q_0)^2, \quad a^2 D^2_0 \geq V^2_0. \quad (B.1)$$

  and

  $$\|H^{-1}_0 A_0 D^{-1}_0\|_\infty =: \rho < 1. \quad (B.2)$$

Using the matrix $D \in \mathbb{R}^{p^* \times p^*}$ and the central point $\upsilon^0 \in \mathbb{R}^{p^*}$ we define the local set $\Upsilon_0(\tau)$ with some fixed value $\tau > 0$.

$$\Upsilon_0(\tau) \overset{\text{def}}{=} \{ \upsilon = (\theta, \eta) \in \Upsilon, \|D_0(\upsilon - \upsilon^0)\| \leq \tau \}.$$

The local conditions only describe the properties of the process $L(\upsilon)$ for $\upsilon \in \Upsilon_0(\tau)$ with some fixed value $\tau > 0$. The global conditions have to be fulfilled on the whole $\Upsilon$. We start with the local conditions.

- **(L_0)** For each $\tau \leq \tau_0$, there is a constant $\delta(\tau)$ such that

  $$\|D_0^{-1}_0 \nabla^2 \mathbb{E} L(\upsilon) D_0^{-1}_0 - I_{p^*}\| \leq \delta(\tau).$$

- **(ED_0)** There exist constants $\nu_{\tau} > 0$ and $g > 0$ such that for all $|\mu| \leq g$;

  $$\sup_{\gamma \in \mathbb{R}^{p^*+m}} \log \mathbb{E} \exp \left\{ \frac{\|\nabla \xi(\upsilon^0), \gamma\|}{g_0 \gamma \|V_0 \gamma\|} \right\} \leq \frac{\nu_0^2 \mu^2}{2}.$$

**Remark B.2.** The matrix $V^2_0$ describes the variability of the process $L(\upsilon)$ around the true point $\upsilon^*$ and in many situations can be set as

$$V^2_0 \overset{\text{def}}{=} \text{Var}\{ \nabla L(\upsilon^*) \}.$$ 

- **(ED_1)** For all $0 < \tau < \tau_0$, there exists a constant $\omega \leq 1/2$ such that for all $|\mu| \leq g$:

  $$\sup_{\upsilon, \upsilon' \in \Upsilon_0(\tau)} \sup_{\|\gamma\| \leq 1} \log \mathbb{E} \exp \left\{ \frac{\mu \gamma^\top D_0^{-1}_0 (\nabla \xi(\upsilon) - \nabla \xi(\upsilon'))}{\omega \|D_0(\upsilon - \upsilon')\|} \right\} \leq \frac{\nu_0^2 \mu^2}{2}.$$
The global conditions are:

\( (\mathcal{L} r) \) For any \( r > r_0 \) there exists a value \( b(r) > 0 \), such that

\[
-IE\mathcal{L}(v, v^\circ) \geq b(r).
\]

\( (\mathcal{E} r) \) For any \( r \geq r_0 \) there exists a constant \( \nu_r > 0 \) and a constant \( g(r) > 0 \) such that

\[
\sup_{v \in \mathcal{V}(r)} \sup_{\mu \leq g(r)} \sup_{\gamma \in \mathbb{R}^p} \log IE \exp \left\{ \mu \langle \nabla \zeta(v), \gamma \rangle \right\} \leq \nu_r \sqrt{m^2 \mu^2 / 2}.
\]

In the sieve approach these conditions have to be satisfied for every \( m \geq m_0 \) for some \( m_0 \in \mathbb{N} \). We also need two condition for the full infinite dimensional functional, where gradients are to be understood in the Fréchet sense. The operator \( \mathcal{D}^2 \) is defined as

\[
\mathcal{D}^2 \overset{\text{def}}{=} -\nabla^2 IE\mathcal{L}(v^*),
\]

where

\[
v^* := \arg\max_v IE[\mathcal{L}(v)] = \arg\max_{(\theta, \eta) \in \mathbb{R}^p \times \mathcal{S}} IE \left[ \mathcal{L} \left( \theta, \sum_{k=1}^{\infty} \eta_k e_k \right) \right].
\]

\( (\mathcal{L} r_{\infty}) \) For any \( r > r_0 \) there exists a value \( b(r) > 0 \), such that

\[
-IE\mathcal{L}(v, v^*) \geq b(r).
\]
References

Andresen, A. (2014). A result on the bias of sieve profile estimation. arXiv:1406.4045

Andresen, A. and Spokoiny, V. (2014). Critical dimension in profile semiparametric estimation. arXiv:1303.4640.

Daubechies, I. (1992). Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics.

Delecroix., M., Haerdle, W., and Hristache, M. (1997). Efficient estimation in single-index regression. Technical report, SFB 373, Humboldt Univ. Berlin.

Haerdle, W., Hall, P., and Ichimura, H. (1993). Optimal smoothing in single-index models. Ann. Statist., 21:157–178.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. J Econometrics, 58:71–120.

Koltchinskii, V. (2012). A remark on low rank matrix recovery and noncommutative bernstein inequalities. IMS Collections, From Probability to Statistics and Back, 9:213–226.

Kosorok, M. (2005). Introduction to Empirical Processes and Semiparametric Inference. Springer in Statistics.

McAleer, M. and da Veig, B. (2008). Single-index and portfolio models for forecasting value-at-risk thresholds. Journal of Forecasting, 27:217–235.

Mendelson, S. (2014). Learning without concentration. arXiv:1401.0304.

Spokoiny, V. (2012). Parametric estimation. Finite sample theory. Ann. Statist., 40(6):2877–2909.

van der Vaart, A. and Wellner, J. (1996). Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics. Springer.