Fisher Information in Group-Type Models

Peter Ruckdeschel

Fraunhofer ITWM, Abt. Finanzmathematik, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
and TU Kaiserslautern, AG Statistik, FB. Mathematik, P.O.Box 3049, 67653 Kaiserslautern, Germany

Abstract

In proofs of $L_2$-differentiability, Lebesgue densities of a central distribution are often assumed right from the beginning. Generalizing Huber (1981, Theorem 4.2), we show that in the class of smooth parametric group models these densities are in fact consequences of a finite Fisher information of the model, provided a suitable representation of the latter is used. The proof uses the notions of absolute continuity in $k$ dimensions and weak differentiability.

As examples to which this theorem applies, we spell out a number of models including a correlation model and the general multivariate location and scale model.

As a consequence of this approach, we show that in the (multivariate) location scale model, finiteness of Fisher information as defined here is in fact equivalent to $L_2$-differentiability and to a log-likelihood expansion giving local asymptotic normality of the model.

Parallel to Huber’s proofs for existence and uniqueness of a minimizer of Fisher information to our situation, we get existence of a minimizer in any weakly closed set $\mathcal{F}$ of central distributions $F$. If, additionally to analogue assumptions to those of Huber (1981), a certain identifiability condition for the transformation holds, we obtain uniqueness of the minimizer. This identifiability condition is satisfied in the multivariate location scale model.

Keywords: Fisher information, group models, multivariate location and scale model, correlation estimation, minimum Fisher information, absolute continuity, weak differentiability, LAN, $L_2$ differentiability, smoothness;
2000 MSC: 62H12, 62F12, 62F35

1. Introduction

1.1. Motivation

$L_2$-differentiability as introduced by LeCam and Hájek appears to be the most suitable setup in which to derive such key properties as local asymptotic normality (LAN) in local asymptotic parametric statistics. In order to show this $L_2$-differentiability however, Lebesgue densities of a central distribution are frequently assumed right from the beginning. In this paper, we generalize Huber (1981, Theorem 4.2) from one-dimensional location to a large class of parametric models, where these Lebesgue densities are in fact a consequence of a finite Fisher information of the
model, provided a suitable definition of the latter is used. This definition may then serve—
again as in [Huber (1981)]—as starting point for minimizing Fisher information along suitable
neighborhoods of the model.

The framework in which this generalization holds covers smooth parametric group models as to
be found in [Bickel et al. (1998)], but is valid even in a somewhat more general setting: The idea
is to link transformations in the parameter space to transformations in the observation space.
The new definition of Fisher information then simply amounts to transferring differentiation
in the parameter space to differentiation—in a weak sense—in the observation space. This is
actually done much in a Sobolev spirit, working with generalized derivatives.

1.2. Organization of the Paper

After an introduction to the setup of smooth parametric group models, in section 2 we list
the smoothness requirements for the transformations and some notation needed for our theorem.
Before stating this theorem, in section 3 we first give a number of examples to which this theorem
applies, the most general of which is the multivariate location and scale model from Example 3.7.
Section 4 provides the main result, Theorem 4.4. In section 5, we spell out the resulting Fisher
information in the examples of section 3. As announced in the motivation, in section 6, culmi-
nating in Proposition 6.2, we show that in the (multivariate) location-scale model finiteness of
Fisher information is equivalent to \( L^2 \)-differentiability as well as to a LAN property. Finally, in
section 7 we generalize Huber’s proofs for existence and uniqueness of a minimizer of Fisher
information to our situation. The proofs are gathered in appendix section Appendix B; The proof
of Theorem 4.4 makes use of the notions of absolute continuity in \( k \) dimensions and of weak
differentiability. Both are provided in an appendix in section Appendix A.

Remark 1.1. The one-dimensional scale model, a particular case of what is covered by this paper, has
been spelt out separately, in a small joint paper with Helmut Rieder, cf. [Ruckdeschel and Rieder (2010)].
with $P_{\theta,k}$ the marginal of $X_k$ and $P_{\theta,j+1,k}$ a regular conditional distribution of $X_j$, given
$X_{j+1} = x_{j+1}, \ldots, X_k = x_k$. In the sequel, we write $y_{i:j}$ for the vector $(y_i, \ldots, y_j)^T$. For a measure
$G$ on $\mathcal{M}(\mathbb{R}^k)$ and a set of indices $J$ we write $G_I$ to denote the joint marginal of $G$ for
coordinates $i \in I$. For $y \in \mathbb{R}^k$ define $y_{-i} := y_{1:1,i+1:k}$, and for $y \in \mathbb{R}^{k-1}$ and $x \in \mathbb{R}$ define the
expression $(x:y)_i := (y_{1:i-1}, x, y_{i+1:k})^T \in \mathbb{R}^k$.

2.2. Model Definition

For a fixed central distribution $F$ on $\mathbb{R}^k$, we consider a statistical model $\mathcal{D} \subset \mathcal{M}(\mathbb{R}^k)$
generated by a family $\mathcal{G}$ of diffeomorphisms $\tau : \mathbb{R}^k \to \mathbb{R}^k$ defined on the observation space.
Denote the inverse of $\tau$ by $\tau = \tau^{-1}$. This family is parametrized by a $p$ dimensional parameter $\theta$,
stemming from an open parameter set $\Theta \subset \mathbb{R}^p$, and this induces the parametric model

$$\mathcal{D} = \{P_\theta | P_\theta = \tau_\theta(F), \ \theta \in \Theta\} \quad (2.2)$$

where $\tau_\theta(F)$ denotes the image measure under $\tau_\theta, F \circ \tau_\theta$.

Remark 2.1. In most examples, $\mathcal{G}$ will be a group, which is also the formulation used in [Lehmann 1983, section
1.3] and [Bickel et al. 1998 Ch. 4]. These authors did not intend to generalize Fisher information, though, and Example
3.3 shows that for our purposes a group structure of for the set $\mathcal{G}$ is not necessary.

2.3. A Smooth Compactification of $\mathbb{R}^k$
For reasons explained in Remark 4.1 we introduce the following compactification $\bar{\mathbb{R}}^k$ of $\mathbb{R}^k$:

Definition 2.2. Let $\mathcal{D}^i([0,1]^k, \mathbb{R})$, $i \in \mathbb{N} \cup \infty$ the space of all continuous real-valued functions
on the domain $[0,1]^k$ which are differentiable $1$ times / arbitrarily often in $(0,1)^k$, and with existing one-sided
derivatives on $\partial[0,1]^k$. We identify this space with functions on $\bar{\mathbb{R}}^k$, using the isometry $\ell$,

$$\ell : [-\infty, \infty]^k \to [0,1]^k, \quad [\ell((x_j))]_i = [e^{x_i}/(e^{x_i} + 1)], \quad (2.3)$$

i.e. let

$$\mathcal{D}^i(\bar{\mathbb{R}}^k, \mathbb{R}) := \mathcal{D}^i([0,1]^k, \mathbb{R}) \circ \ell = \{\varphi \mid \varphi = \psi \circ \ell, \exists \psi \in \mathcal{D}^i([0,1]^k, \mathbb{R})\} \quad (2.4)$$

For later purposes we also note the inverse of $\ell$

$$\kappa : [0,1]^k \to [-\infty, \infty], \quad \kappa(y_1, \ldots, y_k) = (\log(y_j/(1-y_j)))_{j=1,\ldots,k} \quad (2.5)$$

In the same manner, unbounded, continuous functions are defined and denoted by $\mathcal{D}^\infty(\bar{\mathbb{R}}^k, \mathbb{R}^\infty)$.

Remark 2.3. (a) With this definition, $\bar{\mathbb{R}}^k$ becomes a compact metric space.
(b) Integrations along $\bar{\mathbb{R}}^k$ are understood as lifted onto $[0,1]^k$ by $\ell$, i.e. $\int_{\bar{\mathbb{R}}^k} f \, dP = \int_{[0,1]^k} f \circ \kappa d\ell \circ P$.
(c) The choice of $\ell$ resp. $\kappa$ is arbitrary to some extent, but satisfactory for our needs; in fact, we
only have to impose $\ell \in \mathcal{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R}^k)$, $\lim_{x \to -\infty}(\ell(x,y_j)) = 0$, $\lim_{x \to -\infty}(\ell(x,y_j)) = 1$, for each $y \in \mathbb{R}^{k-1}$, $\ell$ strictly isotone in each coordinate, $|\ell'(x)D(\tau_\theta(x))| \leq L_2(F)$, or, for uniformity in $\mathcal{M}(\mathbb{R}^k)$,

$$\sup_h |\ell'(x)D(\tau_\theta(x))| < \infty.$$

(d) For every $\varphi \in \mathcal{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$, the limits $\lim_{x \to \pm \infty} \varphi(x)$ exist and $\lim_{x \to \pm \infty} \frac{d}{dx} \varphi(x) = 0$ for $l \geq 0$, as
is easily seen using the chain rule and by the fact that each summand arising in a derivative has at least a
factor decaying as $\exp(-|x|)$. This also implies that there are functions $\varphi : \mathbb{R} \to \mathbb{R}$ which do not lie in
$\mathcal{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$ but which are in $\mathcal{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$, have existing $\lim_{x \to \pm \infty} \varphi(x)$, and for which $\lim_{x \to \pm \infty} \frac{d}{dx} \varphi(x) = 0$
for $k \geq 0$. Take $1/(x^2 + 1)$, which has no exponentially decaying derivatives.
(e) Consequently, for all \( \varphi \in \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \), \( \int b|\varphi'| d\lambda \) is finite for any bounded, measurable function \( b \) and \( \lim_{x \to \infty} \varphi(x)'/|x|^k = 0 \) for all \( k \in \mathbb{N} \), hence in particular is in \( L_\infty(P) \) for every probability \( P \) on \( \mathbb{B}^k \).

(f) If we allow for mass \( \ell_0P \) in \([0, 1]^k \setminus (0, 1)^k\) — corresponding to measures in \( \mathcal{M}_1(\mathbb{B}^k) \) — the class \( \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \) of compactly supported functions in \( \mathcal{C}^k(\mathbb{R}^k) \) cannot distinguish any measures \( P_1 \neq P_2 \) on \( \mathbb{B}^k \) coinciding on \( \mathbb{B}^k \), whereas \( \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \) is measure determining on \( \mathbb{B}^k \).

(g) The measures \( P_0 \) arising in our model from subsection 2.2 are understood as members of \( \mathcal{M}_1(\mathbb{B}^k) \), defining \( P_0(A) = P_0(A \cap \mathbb{B}^k) \) for \( A \in \mathbb{B}^k \).

2.4. Assumptions

Throughout this paper, we make the following set of assumptions concerning the transformations \( \tau \), which are needed to link differentiation w.r.t. \( \theta \) to differentiation w.r.t. \( x \):

(I) \( P_0 = P_0 \) \iff \( \theta_1 = \theta_2 \).

(D) \( \theta \mapsto t_\theta(x) \) is differentiable with derivative \( \partial_\theta t_\theta(x) \).

(Dk) If \( k > 1 \), \( x \mapsto t_\theta(x) \) is twice differentiable with second derivative \( \partial_\theta^2 t_\theta(x) \).

(C1) If \( k = 1 \), \( x \mapsto D \) in \( \mathcal{C}^1(\mathbb{R}, \mathbb{R}^p) \) and \( x \mapsto e^{-|x|} D \circ \tau_\theta(x) \) is in \( L_2^p(F) \) with

\[
D = D_\theta^{(1)}(x) = \partial_\theta t_\theta(x)/\partial_\theta t_\theta(x) \tag{2.6}
\]

(Ck) If \( k > 1 \), \( x \mapsto D \) in \( \mathcal{C}^1(\mathbb{R}^k, \mathbb{R}^{k \times p}) \), \( x \mapsto e^{-|x|} D \circ \tau_\theta(x) \) is in \( L_2^k(F) \) and \( x \mapsto V \circ \tau_\theta(x) \) is in \( L_2^p(F) \) with

\[
J = (J_\theta(x))_{i,j=1,...,k} = ((\partial_\theta t_\theta)^{-1})_{i,j}(x) \tag{2.7}
\]

\[
D = (D_\theta^{(k)}(x))_{i,j=1,...,k} = [J_\theta^T \partial_\theta t_\theta(T)]_{i,j}(x) \tag{2.8}
\]

\[
V = (V_\theta(x))_{j=1,...,p} = \sum_{i=1}^k \frac{[\partial_i \det(\partial_\theta t_\theta(T))_i - \partial_\theta \det(\partial_\theta t_\theta(T))_i]}{|\det(\partial_\theta t_\theta(T))_i|}(x) \tag{2.9}
\]

Remark 2.4. Using \( \frac{\partial}{\partial A_{ij}} \det A = (A^{-1})_{ij} \det A \) and \( \frac{\partial}{\partial A_{ij}} (A^{-1})_{ij} = (A^{-1})_{ij} (A^{-1})_{il} \) and the chain rule of differentiation, one can show

\[
V_j = |(\det(\partial_\theta t_\theta(T))^{-1}| \sum_{i=1}^k \left[ \delta_{i,j} (\det(\partial_\theta t_\theta(T))_{i,j}) - \partial_\theta \det(\partial_\theta t_\theta(T))_{i,j} \right] = \sum_{i,j,m=1}^k \left( J_{ij,m} - J_{il,m} \right) \partial_\theta^{2} \theta_{i,m} \partial_\theta^2 \theta_{j,l} \tag{2.10}
\]

which motivates requirement (Dk).

In the sequel we use these abbreviations:

Notation 2.5. The set \( \{D = 0\} \) is denoted by \( K \). With \( e_i \) the \( i \)-th canonical unit vector in \( \mathbb{R}^k \) and some \( a \in \mathbb{R}^p \) and \( y \in \mathbb{R}^{k-1} \), define

\[
V_a := V^a, \quad D_a := Da, \quad D_{a:j} := e_i^a D_a, \quad K_i := \{e_i^a D = 0\} \tag{2.11}
\]

Also, for later purposes—c.f. (4.4)—we introduce the functions

\[
\tilde{D} := (D_\theta^{(k)}(x))_{j=1,...,k} = \partial_\theta^2 \theta(T), \quad \tilde{V} = (\tilde{V}_\theta(x))_{j=1,...,p} = V_\theta \tag{2.12}
\]

\[
\tilde{V}_a := \tilde{V}^a, \quad \tilde{D}_a := Da, \quad \tilde{D}_{a:j} := e_i^a \tilde{D}_a, \quad \tilde{K}_i := \{e_i^a \tilde{D} = 0\} \tag{2.13}
\]

Finally, if \( F \ll k \), we write \( f_\theta \) for \( f \circ t_\theta \), with \( f \) a \( k \)-density of \( F \).
We also introduce the following decomposition of $P_\theta$:

$$P_\theta := P^{(0)}_\theta + P^{(0)}_\theta, \quad P^{(0)}_\theta(\cdot) := P_\theta(\cdot \cap K). \quad (2.14)$$

3. Examples

For the following seven popular examples we spell out the transformations $\tau_\theta(x)$ and the respective parameter space and verify the assumptions from the preceding section.

**Example 3.1 (one-dim. location).** $\tau_\theta(x) := x + \theta$, $\theta \in \Theta_1 \subset \mathbb{R}$, $p = k = 1$.

For each $\theta \in \Theta_1$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) and (C1) are satisfied—$\partial_\theta x = 1$, $\partial_\theta t_\theta = 0$, $D(x) = -x$, $K = \emptyset$—any observation $x$ is informative for this problem.

**Example 3.2 (k-dim. location, $k > 1$).** $\tau_\theta(x) := x + \theta$, $\theta \in \Theta_2 \subset \mathbb{R}^k$, $p = k$.

For each $\theta \in \Theta_2$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are satisfied—$\partial_\theta x = 1$, $\partial_\theta t_\theta = -x/\theta$, $D(x) = -x/\theta. \quad \therefore K = \{0\}$, hence the point $x = 0$ is not informative for this problem, and any $x \neq 0$ is.

**Example 3.3 (one-dim. scale).** $\tau_\theta(x) := \theta x$, $\theta \in \Theta_3 \subset \mathbb{R}_{>0}$, $p = k = 1$.

For each $\theta \in \Theta_3$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) and (C1) are satisfied—consider $\partial_\theta t_\theta = -\theta$, $\partial_\theta t_\theta = 0$, $D(x) = 0, K = \emptyset$—any observation $x$ carries information for this problem.

**Example 3.4 (one-dim. loc. and scale).** $\tau_\theta(x) := \theta_2 x + \theta_1$, $\theta \in \Theta_4 = \Theta_1 \times \Theta_3$, $k = 1$, $p = 2$.

For each $\theta \in \Theta_4$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) and (C1) are satisfied—consider $\partial_\theta t_\theta = -\theta_2$, $\partial_\theta t_\theta = 0$, $D(x) = 0, K = \emptyset$—any observation $x$ carries information for this problem.

**Example 3.5 (correlation, $k = 2$; $p = 1$).** To $\sigma_1, \sigma_2 > 0$ known let $\theta \in \Theta_5 = (-1;1)$

$$\tau_\theta : \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto \tau_\theta(x) := J_\theta x,$n \quad J_\theta = \begin{pmatrix} \sigma_1(1 - \theta^2)^{-\frac{1}{2}} & \theta \sigma_1 \\ 0 & \sigma_2 \end{pmatrix} x; \quad (3.1)$$

In contrast to all other examples considered here, this family does not form a group; this may easily be seen, as $J_\theta^{-1}$ does not admit a representation according to (3.1). For each $\theta \in \Theta_5$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are clearly satisfied—$\partial_\theta t_\theta = 0$, $V = 0$, $\partial_\theta t_\theta = (1 - \theta^2)^{-\frac{1}{2}}(x_1 \theta \sigma^{-1}_1 - x_2 \sigma^{-1}_2, 0)^\top$,

$$\partial_\theta t_\theta = (1 - \theta^2)^{-\frac{1}{2}} \begin{pmatrix} \sigma^{-1}_1 & -\theta \sigma^{-1}_2 \\ 0 & (1 - \theta^2)/\sigma_2 \end{pmatrix} .$$

$D = ([x_1 - \sigma_1 \sigma_2]/(1 - \theta^2), 0)^\top$. As $K = \{x \in \mathbb{R}^2 \mid \exists \rho \in \mathbb{R} : x = \rho(\sigma_1; \sigma_2)^\top\}$, $P(K) < 1$ holds, as long as supp($P_\theta$) is not contained in the line $\{\rho(1; \theta \sigma_2/\sigma_1)^\top, \rho \in \mathbb{R}\}$, or equivalently, as long as supp($F$) $\not\subset \{\rho((1 - \theta^2)^{-\frac{1}{2}}; \theta)^\top, \rho \in \mathbb{R}\}$.

**Example 3.6 (k-dim. scale, $k > 1$).** $\tau_\theta(x) := \theta x$, defined for $\Theta_6 = \{S \in \mathbb{R}^{k \times k} \mid S = S^\top > 0\}$, $p = (k^2 - 1)/2$. The symmetry restriction is imposed on $\mathbb{R}^{k \times k}$, allowing only for symmetric variations in the parameter.
Again, for each \( \theta \in \Theta_k \), \( \tau_\theta(\cdot) \) is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are satisfied—\( \partial_\theta^2 t_\theta = 0 \), \( \partial_\theta t_\theta = (\theta^{-1})_{i,i} \),
\[
\begin{align*}
\partial_\theta_{i,j} t_{\theta,i} &= -\frac{1}{2}[(\theta^{-1})_{i,i}(\theta^{-1}x)_i]_j + (\theta^{-1})_{i,j} (\theta^{-1}x)_i, \\
D &= D_{i,j} = \frac{1}{2}[(\frac{1}{2} \otimes \theta^{-1}x)_{i,i} + (\frac{1}{2} \otimes \theta^{-1}x)_{i,j}], \quad V = 0.
\end{align*}
\]

For each symmetric matrix \( a \in \text{GL}(k) \), we have \( D(x)a = \theta^{-1}a\theta^{-1}x;\ K = \{0\} \)—any observation \( x \neq 0 \) carries information for this problem.

**Example 3.7** \((k\text{-dim. location and scale, } k > 1)\). \( \tau_\theta(x) := \theta_2 x + \theta_1 \), for \( \theta \in \Theta_k = \mathbb{R}^k \times \Theta_6 \), \( p = k + (k+1)/2 = k(k+3)/2 \).

For each \( \theta \in \Theta_k \), \( \tau_\theta(\cdot) \) is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are satisfied—\( \partial_\theta^2 t_\theta = 0 \), \( \partial_\theta t_\theta = \theta_2^{-1} \); splitting off the indices for the parametric dimensions into the location part [a single index] and the scale part [a double index], we get
\[
\begin{align*}
\partial_\theta_{i,j} t_{\theta,i} &= -\theta_2^{-1}_{i,i}, \\
\partial_\theta_{i,j} t_{\theta,i} &= -\frac{1}{2}[(\theta^{-1}_2)_{i,i}(\theta_2^{-1}(x - \theta_1))_i]_j + (\theta^{-1}_2)_{i,j} (\theta_2^{-1}(x - \theta_1))_i; \\
D_{i,i} &= 0, \\
D_{i,j} &= \frac{1}{2}[(\frac{1}{2} \otimes \theta_2^{-1}(x - \theta_1))_{i,i} + (\frac{1}{2} \otimes \theta_2^{-1}(x - \theta_1))_{i,j}].
\end{align*}
\]

Just as in Example 3.2, any observation \( x \) carries information for this problem.

### 4. Main Theorem

In [Huber, 1981], Definition 4.1 and Theorem 4.2, we find a result on the Fisher information in the one dimensional location case which is central for the famous minimax M estimator result of [Huber, 1964]. The idea is to express Fisher information as a supremum, i.e.
\[
\mathcal{F}(F) := \sup \left\{ \left( \frac{\int \varphi' dF}{\int \varphi^2 dF} \right)^2 \bigg| \varphi \neq 0 \in \mathcal{D}_1 \right\}. \tag{4.1}
\]

With this definition, [Huber, 1981] Thm 4.2) achieves a representation of Fisher information without assuming densities of the central distribution: \( \mathcal{F}(F) \) is finite iff \( F \) is a.c. with a.c. Lebesgue density \( f \) such that \( \int (f'/f)^2 f dx < \infty \), which in this case is just \( \mathcal{F}(F) \).

**Remark 4.1.**

(a) The proof in [Huber, 1981] is credited to T. Liggett and is based on Sobolev-type ideas; we take these up to generalize the result to more general models and higher dimensions.

(b) The set \( \mathcal{D}_1 \) in (4.1) plays the role of a set of test functions as in the theory generalized functions, compare [Rudin, 1991] Ch. 6). In the cited reference, Huber uses \( \mathcal{D}_1 = \mathcal{C}_c^1(\mathbb{R}, \mathbb{R}) \), the subset of compactly supported functions in \( \mathcal{C}^1(\mathbb{R}, \mathbb{R}) \). In the proof later, we will need that the sets
\[
\mathcal{D}_{D_{i,j}} := \{ D_{i,j} \partial_\theta \varphi \mid \varphi \in \mathcal{D}_k, a \in \mathbb{R}^p \}
\]
are dense in \( L_2(P_{\theta}^{(1)}) \). Contrary to the one-dimensional location case, for \( \mathcal{D}_k = \mathcal{C}_c^1(\mathbb{R}^k, \mathbb{R}) \) and general \( D_{i,j} \), we did not succeed to prove this; nor can we work with \( \mathcal{D}_k = \mathcal{C}_c^1(\mathbb{R}^k, \mathbb{R}) \), the set of continuously
differentiable functions with compactly supported derivatives, as used for the one dimensional scale model in [Rückdeschel and Rieder (2010) Lem. A.1]: The crucial approximation of the constant function 1 by functions $\phi \in C_c(\mathbb{R}^k, \mathbb{R})$, with $|\phi| \leq 1$, $|D_{a,j}\phi| \leq 1$, and $|D_{a,j}\phi| \to 0$ pointwise, fails for functions $D_{a,j}$ growing faster than $|x|$ for large $|x|$. Hence, instead we use the larger set $C_{cr}(\mathbb{R}^k, \mathbb{R})$ from Definition 2.3.

Definition 4.2. In model $\mathcal{P}$ from (2.2), assume assume (I) and (D). Let $a \in \mathbb{R}^p$, $|a| = 1$.

$k = 1$: Assume (C1). Let $\mathcal{D}_1 = C(\mathbb{R}^1, \mathbb{R})$, $D$ from (2.6). Then for $\theta \in \Theta$ we define

$$\mathcal{I}_\theta(F; a) := \sup \left\{ \frac{\left(\int |\varphi^\top D_{a} | dP_{\theta} \right)^2}{\int \varphi^2 dP_{\theta}} \mid 0 \neq \varphi \in \mathcal{D}_1 \right\}, \quad (4.2)$$

$k > 1$: Assume (Dk) and (Ck). Let $\mathcal{D}_k = C(\mathbb{R}^k, \mathbb{R})$, $D$ and $V$ from (2.8) and (2.9). Then for $\theta \in \Theta$ we define

$$\mathcal{I}_\theta(F; a) := \sup \left\{ \frac{\left(\int |\nabla \varphi^\top D_{a} + \varphi V_{a} | dP_{\theta} \right)^2}{\int \varphi^2 dP_{\theta}} \mid 0 \neq \varphi \in \mathcal{D}_k \right\}, \quad (4.3)$$

Remark 4.3. (a) As $\tau_\theta$, resp. $t_\theta$ map $\mathcal{D}_k$ onto itself, we may use the identification $\psi = \varphi \circ \tau_\theta$ to see that by the transformation formula

$$\mathcal{I}_\theta(F; a) := \sup \left\{ \frac{\left(\int |\nabla \psi^\top D_{a} + \psi V_{a} | dF \right)^2}{\int \psi^2 dF} \mid 0 \neq \psi \in \mathcal{D}_k \right\}, \quad (4.4)$$

(b) In particular, the transformation formula $\int \rho(x) P_\theta(dx) = \int \rho \circ \tau_\theta dF$, entails that except for the correlation model of Example 3.5, finiteness of the Fisher information for one $\theta \in \Theta$ implies finiteness for every $\theta \in \Theta$: Indeed, considering $D_{k}^{(\psi)} \circ \tau_\theta$ in all these models, we see that in every case, $D_{k}^{(\psi)} \circ \tau_\theta = D_{k}^{(\psi)}$, where we write $id$ referring to the parameter-value $\theta$ yielding $t_\theta = id$, while at the same time $V = 0$. So in fact we could define the Fisher information of $F$ for one reference parameter, and its finiteness then entails finiteness in the whole parametric model.

(c) In general, finiteness will however depend on the actual parameter value, which is why we define Fisher information at $F$ with reference to $\theta$, notationally transparent as $\mathcal{I}_\theta(F; a)$.

With Definition 4.2 we generalize Huber (1981 Thm. 4.2) to

Theorem 4.4. In model $\mathcal{P}$ from (2.2) assume that for some fixed $\theta \in \Theta$, (I), and, if $k = 1$, (D), and (C1), resp., if $k > 1$, (Dk) and (Ck) hold. Then (the sets of) statements (i) and (ii) are equivalent:

(i) $\sup_{|x|=1} \mathcal{I}_\theta(F; a) < \infty$

(ii) (a) $F$ admits a $\lambda^k$ density $f$ on $t_\theta(K^\circ)$.

(b) For every $a \in \mathbb{R}^p$, and $i = 1, \ldots, k$

$$\lim_{|x| \to 0} \left[ f_\theta \det \partial_a t_\theta \left| D_{a,i} \right|(x; y) \right] = 0$$

(c) For every $a \in \mathbb{R}^p$, and $i = 1, \ldots, k$ $f_\theta |\det \partial_a t_\theta | D_{a,i}$ is a.c. in $k$ dimensions in the sense of Definition A.3.
Example 5.4 (one-dim. location and scale). If (i) resp. (ii) holds, \( I \) and \( I \) respectively
of Theorem 4.4, whenever the map \( \Lambda \) is weaker than ours, so formally our implication (ii) is harder, (i) \( \Rightarrow \) (ii) is easier than his.

Example 5.1 (one-dim. location). \( \Lambda_0(x) := \Lambda_f(x - \theta) \), \( \mathcal{I}_0 = \int \Lambda_f dF \). The supremal definition of \( \mathcal{I}(F) \) is (4.1), but with \( \mathcal{D}_1 = \mathcal{D}_0 \) instead of \( \mathcal{D}_1 = \mathcal{D}_0 \).

Example 5.2 (k-dim. location, \( k > 1 \)). \( \Lambda_0(x) := \Lambda_f(x - \theta) \), \( \mathcal{I}_0 = \int \Lambda_f \Lambda_f dF \), \( \mathcal{I}_0(F; a) = a^T \mathcal{I}_0 a \). The supremal definition of \( \mathcal{I}(F) \) is

\[
\mathcal{I}_0(F) := \sup \left\{ \frac{\left( \int f \phi^a x dF(x) \right)^2}{\int \phi^a dF} \bigg| \phi \neq 0 \in \mathcal{D}_1 \right\}
\]

Example 5.3 (one-dim. scale). \( \Lambda_0(x) := \frac{1}{\theta} \left[ (x/\theta) \Lambda_f(x/\theta) + 1 \right] \), \( \mathcal{I}_0 = \frac{1}{\theta^2} \mathcal{D}_1 = \frac{1}{\theta^2} \int (x \Lambda_f - 1)^2 dF = \frac{1}{\theta^2} \left( \int x^2 \Lambda_f^2 dF - 1 \right) \). The supremal definition of \( \mathcal{I}(F) \) is

\[
\mathcal{I}_1(F) := \sup \left\{ \frac{\left( \int f \phi^a x F(dx) \right)^2}{\int \phi^a dF} \bigg| \phi \neq 0 \in \mathcal{D}_1 \right\}
\]

Example 5.4 (one-dim. location and scale).

\[
\Lambda_0(x) := \frac{1}{\theta_2} \left( \Lambda_f \frac{x - \theta_1}{\theta_2}, \frac{x - \theta_1}{\theta_2} \Lambda_f \left( \frac{x - \theta_1}{\theta_2} \right) + 1 \right) \tau,
\]

\[
\mathcal{I}_0(x) := \frac{1}{\theta_2} \mathcal{I}_{0;1}(x) = \frac{1}{\theta_2} \left( \int \Lambda_f^2 dF \int x \Lambda_f^2 dF \int (x \Lambda_f - 1)^2 dF \right),
\]

and \( \mathcal{I}_0(F; a) = a^T \mathcal{I}_0 a / \theta_2 \). With \( a = (a_1, a_2)^T \), the supremal definition of \( \mathcal{I}(F) \) is

\[
\mathcal{I}_{0;2}(F; a) := \sup \left\{ \frac{\left( \int f (a_1 + a_2 x) \phi^a x F(dx) \right)^2}{\int \phi^a dF} \bigg| \phi \neq 0 \in \mathcal{D}_1 \right\}
\]
Example 5.5 (correlation, \( k = 2; \ p = 1 \)).

\[ \sigma_2 \sigma_1 \sqrt{1 - \theta^2} p_0(dx_1, dx_2) = f \left( \frac{x_1/\sigma_1 - \theta x_2/\sigma_2}{\sqrt{1 - \theta^2}} \right), \quad \sigma_2^2(x_1, dx_2) = f_2 \left( \frac{x_2}{\sigma_2} \right) \]  

or with \( f = f_{1/2} \)

\[ \sigma_1 \sqrt{1 - \theta^2} p_{0,1/2}(x_1, x_2) = f_{1/2} \left( \frac{x_1/\sigma_1 - \theta x_2/\sigma_2}{\sqrt{1 - \theta^2}} \right), \quad \sigma_2 p_{0,2}(x_1, x_2) = f_2 \left( \frac{x_2}{\sigma_2} \right) \]  

and

\[ (1 - \theta^2) \Lambda_\theta(x_1, x_2) = \frac{f'_{1/2} \left( \frac{x_1/\sigma_1 - \theta x_2/\sigma_2}{\sqrt{1 - \theta^2}} \right)}{\sqrt{1 - \theta^2}} \left( \frac{\theta x_1/\sigma_1 - x_2/\sigma_2}{\sqrt{1 - \theta^2}} \right) + \theta, \quad \mathcal{I}_\theta = \int \Lambda_\theta^2 dP_\theta \]  

The supremaal definition of \( \mathcal{I}(F) \) is

\[ \mathcal{I}_\theta(F) := \sup \left\{ \left( \frac{\int [\theta x_1 - \sqrt{1 - \theta^2} x_2] \left( \partial_\theta F \right)(x_1, x_2) F(dx_1, dx_2) }{(1 - \theta^2)^2 \int \phi^2 dF} \right)^2 \left| \begin{array}{c} \phi = 0 \end{array} \right\} \right\} \]  

Example 5.6 (\( k \)-dim. scale, \( k > 1 \)). We give both vech expressions and matrix expressions, using symmetrized Kronecker products. We start with unsymmetrized versions.

\[ \Lambda_\theta^0(x) = \theta^{-1} \Lambda_{\theta_2}(\theta^{-1} x), \quad \Lambda_{\theta_2}^0(x) = \Lambda_f(x) x^T - \mathbb{I}_k, \quad \Lambda_{\theta_0}^0(x) = \frac{1}{2} \left[ \Lambda_{\theta_2}^0(x) + \Lambda_\theta^0(x)^T \right], \quad \Lambda_\theta^0(x) = \text{vech}[\Lambda_\theta(x)] \]

This can also be written as \( \Lambda_\theta^0(x) = \text{vech}[\theta^{-1} \otimes \Lambda_{\theta_2}(x)] \). In matrix notation this yields

\[ \mathcal{I}_\theta = \left( \left( \theta^{-1} \otimes \theta^{-1} \right) \otimes \int (\Lambda_f x^T - \mathbb{I}_k)^{\otimes 2} dF \right), \]

in vector notation \( \mathcal{I}_\theta = \int \Lambda_\theta^0 (\Lambda_\theta^0)^T dF \). Working with \( a = a^T \in \mathbb{R}^{k \times k} \), we get

\[ \text{vech}(a)^T \Lambda_\theta^0(x) = \Lambda_f (\theta^{-1} x)^T \theta^{-1} a \theta^{-1} x - \text{tr}(\theta^{-1} a), \]

\[ \mathcal{I}_\theta(F, a) = \int \left( \Lambda_f (y)^T \theta^{-1} a y - \text{tr}(\theta^{-1} a) \right)^2 F(dy) \]

For symmetric \( a \), the supremaal definition of \( \mathcal{I}(F) \) is

\[ \mathcal{I}_\theta(F, a) := \left\{ \sup \left( \frac{\int \partial_\theta F(x) \theta^{-1} ax F(dx)}{\int \phi^2 dF} \right)^2 \left| \begin{array}{c} \phi = 0 \end{array} \right\} \right\} \]  

Example 5.7 (\( k \)-dim. location and scale, \( k > 1 \)). Partitioning \( \Lambda \) into a location block \( (1) \) and a scale block \( (s) \), we get

\[ \Lambda_{s,0,0_1}(x) = \theta_2^{-1} \Lambda_{s,0,0_3}(\theta_2^{-1}(x - \theta_1)), \quad \Lambda_{s,0,0_2}(x) = \Lambda_f(x) \]

\[ \Lambda_{s,0,0_1}(x) = \theta_2^{-1} \Lambda_{s,0,0_3}(\theta_2^{-1}(x - \theta_1)), \quad \Lambda_{s,0,0_2}(x) = \Lambda_f(x) x^T - \mathbb{I}_k \]

\[ \Lambda_{s,0,0_1}(x) = (\Lambda_{0,0,0_3}^0(x) + \Lambda_0^0(x)^T)/2, \quad \Lambda_\theta^0(x) = \text{vech}(\Lambda_{s,0,0_1}(x)) \]
should be possible to treat the problem using only one dimensional densities. Factorizations (5.5) and (5.6) as good as knowledge of
with

Consider $Y$ representation. But even then it is of rather marginal interest, as may be seen in the following example:

We have found two further observations $X$ that anyway, not every conditional densities for each fixed $F$ instead of (5.4), we could allow for any second marginal $\theta$ zero of statistical decision problems; if on the other hand, we take

$a = (a_1^T; \text{vech}(a_2^T))^T$, $a_t \in \mathbb{R}^d$, $a_t = a_t^T \in \mathbb{R}^{k \times k}$, we get

$$\text{vech}(a)^T \Lambda_0^T(x) = a_1^T \theta_2^{-1} \Lambda_f(\theta_2^{-1}(x - \theta_1)) + \Lambda_f(\theta_2^{-1}(x - \theta_1))^T \theta_2^{-1} a_1 \theta_2^{-1} (x - \theta_1) - \text{tr}(\theta_2^{-1} a_1),$$

$$\mathcal{J}_a(F, a) = \int (a_1^T \theta_2^{-1} \Lambda_f(y) + \Lambda_f(y)^T \theta_2^{-1} a_1 \theta_2^{-1} a_1 - \text{tr}(\theta_2^{-1} a_1))^2 F(dy)$$

The supremal definition of $\mathcal{J}(F)$ is

$$\mathcal{J}_\theta(F, a) := \sup \left\{ \frac{\int \varphi(x)^T \theta_2^{-1} [a_x x + a_i] F(dx)^2}{\int \varphi^2 dF} \bigg| 0 \neq \varphi \in \mathcal{D}_h \right\} \quad (5.9)$$

To keep the order of the examples as in section 5.3 we place a remark here, concerning Example 5.3.

**Remark 5.8.** The fact that we are dealing with a one dimensional parameter seems to indicate that it should be possible to treat the problem using only one dimensional densities. Factorizations (5.5) and (5.6) seem to point into the same direction, as they seem to suggest that working with

$$\sigma_1 \sqrt{1 - \theta^2} P_\theta(dx_1, dx_2) = f_{1,2} \left( \frac{x_1}{\sigma_1 - \theta x_2 / \sigma_2}, \frac{x_2}{\sigma_2} \right) \lambda(dx_1) F_2(\sigma_2^{-1} dx_2) \quad (5.10)$$

instead of (5.4), we could allow for any second marginal $F_2$—possibly even $F_2 \perp \lambda$ — and just focus on the conditional densities for each fixed $x_2$ section.

Theorem 4.4 however, excludes that possibility for finite Fisher information. To be fair, one has to admit that anyway, not every $F$ with $P_\theta = \int \theta F$ could be allowed for (5.10), but only exactly those achieving this representation. But even then it is of rather marginal interest, as may be seen in the following example:

Consider $Y_1 \sim N(0, 1)$, $Y_2 \sim \pm 1$ with $P(Y_2 = 1) = P(Y_2 = -1) = 1/2$, $Y_1, Y_2$ independent and $F := \mathcal{L}(Y_1, Y_2)$. Then for any $\theta \in \Theta_2$, $X = \tau_0(Y) = ((1 - \theta^2)^{1/4} Y_1 + \theta Y_2, Y_2)$, and recovering $\theta$ from observations of $X$ amounts to estimating $E[X_1|X_2 = x_2]$ for $x_2 = \pm 1$ — a task falling into the usual $Op(n^{-1/2})$-type of statistical decision problems; if on the other hand, we take $F = \mathcal{L}(Y_1, (1 - \alpha^2)^{1/4} Y_1 + \alpha Y_2)$, for any $0 < |\alpha| < 1$, then, for $\theta \neq (1 - \alpha^2)^{1/4}$, $\mathcal{L}(X)$ is concentrated on two lines $X_2 = a_1 + \beta X_1$, $i = 1, 2$ with $\beta = (1 - \alpha^2)^{1/2} / (1 - \alpha^2)^{1/4} + \theta (1 - \alpha^2)^{1/2}$. But as we assume $F$ to be known, knowledge of $\beta$ is just as good as knowledge of $\theta$. Having fixed an observation $X^{(0)}$, $\beta$ may be recovered exactly, as soon as we have found two further observations $X^{(1)}$ and $X^{(2)}$ both lying on the same line as $X^{(0)}$, which will happen in finite time almost surely. Thus here a single observation must have infinite information on $\theta$ — which is just according to our theorem.
6. Consequences for the LAN Approach

In general finiteness of Fisher information does not imply $L_2$-differentiability without additional assumptions like, e.g. that for $\lambda^k$ almost all $x$ and for all $p \in \mathbb{R}^p$ the map $x \mapsto p_{\theta} + s p(x)$ is a.c. and the Fisher information $\mathcal{I}_\theta$ is continuous in $\theta$ — c.f. Le Cam (1986, 17.3 Prop.4).

All examples from section 3—except for the correlation example, Example 3.5—provide more structure, though. They may all be summarized in the (multivariate) location scale model of Example 3.7. First of all, due to the invariance/dilation relations of Lebesgue measure w.r.t. affine transformations, we may limit attention to the reference parameter $(0, I_k)$. Even more though, we have the following generalization of Lemmas by Hajek (1972) (one-dimensional location) and Swensen (1980, Ch.2, Sec.3) to the multivariate location case.

Proposition 6.1. Assume that in the multivariate location and scale model (3.7), Fisher information as defined in (5.8) is finite for some parameter value. Then the model is $L_2$-differentiable for any parameter value.

Hence Theorem 4.4 gives a sufficient condition for these models to be $L_2$-differentiable and as a consequence to be LAN.

On the other hand, $L_2$-differentiability requires finiteness of $\mathcal{I}_\theta$, so that in the multivariate location and scale case, for all central distributions $F$, the model with central distribution $F$ is $L_2$ differentiable iff $\sup_a \mathcal{I}_\theta(F,a) < \infty$.

In the i.i.d. setup Le Cam (1986, 17.3 Prop.2) even show that $L_2$-differentiability is both necessary and sufficient to get an LAN expansion of the likelihoods in form

$$\log dP^n_{\theta+h/\sqrt{n}}/dP^n_\theta = \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \Lambda_\theta(x_i) - \frac{1}{2} h^T \mathcal{I}_\theta h + o_P(n)$$

with some $\Lambda_\theta \in L_2(P_\theta)$ and $0 < \mathcal{I}_\theta = \mathbb{E}[\Lambda_\theta \Lambda_\theta^T] < \infty$, so again in the setup of the (multivariate) location scale model of Example 3.7 finiteness of Fisher information is both necessary and sufficient to such an LAN expansion. Altogether we have

Proposition 6.2. In models 3.1, 3.2, 3.3, 3.4, 3.6, 3.7, the following statements are equivalent

(i) The respective Fisher information from (4.3) is finite for any parameter value.
(ii) Conditions (ii) of Theorem 4.4 hold for any parameter value.
(iii) The model is $L_2$-differentiable for any parameter value.
(iv) The model admits the LAN property (6.1) for any parameter value.

Remark 6.3. The proof uses the translation invariance and the transformation property under dilations of $k$-dimensional Lebesgue measure, so there is not much room for extensions beyond group models induced by subgroups of the general affine group.

7. Minimization of the Fisher information

Representations (4.2) resp. (4.3) for Fisher information allow for minimization, resp. to maximization of the trace or maxev of $\mathcal{I}_\theta$ w.r.t. the central distribution $P_\theta$ or $F$. In this paper, we settle the questions of (strict) convexity and lower continuity just as in Huber (1981), but replace
To this end define for a function and uniqueness of a minimizing \( F(0) \) in some suitable neighborhood of the (ideal) model. To this end define for \( a \in \mathbb{R}^p, \varphi \in \mathcal{D}_k, \| \varphi \circ \tau_0 \|_{L_2(F)} \neq 0 \)

\[
\mathcal{J}_\theta(F; a; \varphi) := \left( \frac{\int \varphi \nabla^2 D_a + \varphi V_a d[\tau_0 F]}{\int \varphi^2 d[\tau_0 F]} \right)^2
\]

(7.1)

and

\[
\tilde{\mathcal{J}}_\theta(F) := \sup \mathcal{J}_\theta(F; a), \quad a \in \mathbb{R}^p, |a| = 1
\]

(7.2)

### 7.1. Weak Lower Semicontinuity and Convexity

To show weak lower semicontinuity and convexity, we use that for fixed \( \varphi \in \mathcal{D}_k, \varphi \neq 0 \|P_0\| \), \( F \mapsto \mathcal{J}_\theta(F; a; \varphi) \) is weak continuous (by definition) and convex (by \cite{huber1981}, Lemma 4.4). Essentially we may then use that the supremum of continuous functions is lower semicontinuous and the supremum of convex functions remains convex; but some subtle additional arguments are needed as the set of \( \varphi \)'s over which we are maximizing may vary from \( F \) to \( F \); these can be found in \cite{ruckdeschel2010}, Proof to Prop. 2.1. Altogether we have shown

**Proposition 7.1.** For each \( a \in \mathbb{R}^p \), the mapping \( F \mapsto \mathcal{J}_\theta(F; a) \) is weakly lower-semicontinuous and convex in \( F \in \mathcal{M}_1(\mathbb{R}^k) \). The same goes for \( F \mapsto \tilde{\mathcal{J}}_\theta(F) \).

**Remark 7.2.** Using \( \mathbb{R}^k \) from Definition 4.2, we work with a compact definition space right away, which moreover is endowed with a separable metric, so any subset of probability measures on \( \mathbb{R}^k \) is tight, hence by Prokhorov’s theorem weakly relatively sequentially compact.

**Corollary 7.3.** In any weakly closed set \( \mathcal{F} \subset \mathcal{M}_1(\mathbb{R}^k) \), both \( \mathcal{J}_\theta \) and \( \tilde{\mathcal{J}}_\theta, a \)—for fixed \( a \)—attain their minimum in some \( F_0 \in \mathcal{F} \).

### 7.2. Strict Convexity—Uniqueness of a Minimizer

We essentially take over the assumptions of \cite{huber1981}, we fix \( \theta \in \Theta \) and consider variations in \( F \) of the following form: For \( F_1 \in \mathcal{M}(\mathbb{R}^p) \), \( t = 0, 1 \) consider

\[
\tilde{F}_t := (1-t)F_0 \circ \theta_t + tF_1 \circ \theta_t
\]

(7.3)

We distinguish cases (a) and (i ), i.e., of a given one-dimensional projection \( a \neq 0 \), and the corresponding maximal eigenvalue, respectively.

**Proposition 7.4.** Under assumptions

(a) The set \( \mathcal{F} \) of admitted central distributions \( F \) is convex.

(b) There is a \( F_0 \in \mathcal{F} \) minimizing

(\( \mathcal{J} \)) \( \mathcal{J}_\theta(F_0; a) \) along \( \mathcal{F} \) and \( \mathcal{J}_\theta(F_0; a) < \infty \).

(\( \tilde{\mathcal{J}} \)) \( \tilde{\mathcal{J}}_\theta(F) \) along \( \mathcal{F} \) and \( \tilde{\mathcal{J}}_\theta(F_0) < \infty \).

(c) The set where the Lebesgue–density \( f_0 \) of \( F_0 \) is strictly positive is convex and contains the support of every \( F_t \) derived from some \( F_t \in \mathcal{F} \).

(d) \( \lambda_1(\{ x|a^2\theta_0(x) = 0 \}) = 0 \)

(\( \tilde{\mathcal{J}} \)) \( \lambda_1(\{ x|\exists a: |a| = 1 \text{ s.t. } a^2\theta_0(x) = 0 \}) = 0 \)
the map $F \mapsto \mathcal{F}_0(F; a)$ (case (a)) resp. $F \mapsto \mathcal{F}_0(F)$ (case (i)) is strictly convex, hence there
is a unique minimizer of $F_0$.

**Remark 7.5.** Assumption (d) holds for the $k$ dimensional location scale model of Example [3.7]. For symmetric
$\alpha_i$ and the scale part $\theta_i$ it holds that $\alpha_i^2 \partial_\alpha \theta_i = \alpha_i \theta_i^{-1} x$. But for any $\alpha_i$ with $|\alpha_i| = 1$,
$\dim \ker \alpha_i \theta_i^{-1} \leq k - 1$, hence $\lambda^k(\ker \alpha_i \theta_i^{-1}) = 0$.

### 7.3. Existence of a Maximizer of $\text{tr} \mathcal{F}_0^{-1}(F)$

**Proposition 7.6.** Let $\mathcal{F}$ be a weakly closed subset of $\mathcal{M}_1(\mathbb{R}^k)$. Assume that for all $0 \neq a \in \mathbb{R}^p$,
$\min_{F \in \mathcal{F}} \mathcal{F}(F; a) > 0$. Then the function $F \mapsto \text{tr} \mathcal{F}_0^{-1}(F)$ is weakly upper-semicontinuous on $\mathcal{F}$,
and consequently, attains its maximum along $\mathcal{F}$ in some $F_0 \in \mathcal{F}$.

### Appendix A. Functional Analysis and Generalized Differentiability

#### Appendix A.1. Dense Functions

**Proposition A.1.** Let $\mu$ be a $\sigma$-finite measure on $\mathbb{B}^k$. Then the set $\mathcal{C}_c^\infty(\mathbb{B}^k, \mathbb{R})$ is dense in any $L_p(\mu)$,
$p \in [1, \infty)$. In particular, there is a $c_0 \in (0, \infty)$ s.t. for any $a < b \in \mathbb{R}$ and any $\delta > 0$ there is a $\varphi = \varphi_{a,b,\delta} \in \mathcal{C}_c^\infty([a, b], [0, 1])$, with $\varphi \equiv 0$ on $[a - \delta; b + \delta]^c$, $\varphi \equiv 1$ on $[a + \delta; b - \delta]$ and $|\varphi| \leq c_0/\delta$.

Proof: Denerness is a consequence of Lusin’s Theorem, compare [Rudin 1974] Thm. 3.14. To achieve the
universal bound $c_0$, we may use functions $\bar{f}(t) = \left( \int_a^t \tilde{f}(s) ds \right) / \left( \int_a^t f(u) du \right)$, for $f(t) = e^{-1/t}$.
$\bar{f}(t)$

#### Appendix A.2. Absolute Continuity

We recall the following characterization of absolute continuity [notation a.c.] of functions $F : \mathbb{R} \to \mathbb{R}$
that can be found in [Rudin 1974] Ch. 8.

**Theorem A.2.** For $F : [a, b] \to \mathbb{R}$, $a < b \in \mathbb{R}$ the following statements 1. to 3. are equivalent

1. $F$ is a.c. on $[a, b]$
2. (a) $F'(x)$ exists $\lambda(dx)$ a.e. on $[a, b]$ and $F' \in L_1(\lambda_{[a,b]})$.
   (b) $F(x) - F(a) = \int_a^x F'(s) \lambda(ds)$ for all $x \in [a, b]$.
3. There is some $u \in L_1(\lambda_{[a,b]})$ s.t. for $x \in [a, b]$, $F(x)$ has the representation

$$F(x) = F(a) + \int_a^x u(s) \lambda(ds)$$

We also recall that a.c. functions, are closed under products [Dudley, R.M. 2002, 7.2 Prob.4]. In particular,
integration by parts is available. In this paper, we call a function $F : \mathbb{R} \to \mathbb{R}$ a.c. if the equivalent statements
1. to 3. from Theorem [A.2] are valid for each compact interval $[a, b] \subset \mathbb{R}$.

### Appendix A.3. Absolute Continuity in Higher Dimensions

A little care has to be taken about null sets when transferring absolute continuity to higher dimensions.
The next definition is drawn from [Simader 2001].

**Definition A.3.** A function $f : (\mathbb{R}^k, \mathbb{B}^k) \to (\mathbb{R}, \mathbb{B})$ is called absolutely continuous (in $k$ dimensions), if for
every $i = 1, \ldots, k$ there is a set $N_i \subset \mathbb{B}^{k-1}$ with $\lambda^{k-1}(N_i) = 0$ s.t. for $y \in N_i^c$, the function $f_i : (\mathbb{R}, \mathbb{B}) \to (\mathbb{R}, \mathbb{B})$, $x \to f_i(x) = f((x:y)_i)$ is a.c. in the usual sense.

In the proof of (ii) $\Rightarrow$ (i) in Theorem [4.4] we need the following lemma:
Lemma A.4. Let \( f : \mathbb{R}^k \to \mathbb{R} \) a.c. in \( k \) dimensions. Then for each \( i = 1 \ldots k \)
\[
\lambda^k(\{ f = 0 \}, \{ \partial_i f \neq 0 \}) = 0
\]  
(A.1)

Proof: Let \( g(z) = 1_{\{f(z) = 0\}}(z) \). Then \( g \geq 0 \) and Tonelli applies, so the section-wise defined function \( h_i(x) := g((x; y_i)) \) is measurable for each \( y \in \mathbb{B}^{k-1} \) and defining the possibly infinite integrals \( H(y) := \int h_i(x) \lambda(dx) \), we get \( \lambda^k(\{ f = 0 \}, \{ \partial_i f \neq 0 \}) = \int g \, d\lambda = \int H(y) \lambda^{k-1}(dy) \). But for each \( y \) the instances \( x \) where \( h_i(x) = 0 \), \( h'_i(x) \neq 0 \) are separated by open one-dim. sets where \( h_i = 0 \), \( h'_i(x) \neq 0 \) implies that for some \( 0 < |x' - x| < \varepsilon \), \( |f(x')| > |x' - x| \lambda'(x)/2 > 0 \). Hence at most there can be a countable number of such \( x \), and thus \( H(y) = 0 \) for each \( y \).

Appendix A.4. Weak Differentiability

For proving absolute continuity in Theorem 4.4 we have worked with the notion of weak differentiability; to this end we compile the following definitions and propositions again drawn from Simader (2001), which have specialized to differentiation of order one.

Definition A.5. Let \( u \in L^1_m(\lambda^k) \), \( 1 \leq i \leq k \). Then \( v_i \in L^1_m(\lambda^k) \) is called weak derivative of \( u \) (with respect to \( x_i \)), denoted by \( \partial_i u \), if
\[
\int_{\mathbb{R}^k} u \partial_i \varphi \, d\lambda^k = -\int_{\mathbb{R}^k} v_i \varphi \, d\lambda^k \quad \forall \varphi \in C^\infty_c(\mathbb{R}^k, \mathbb{R})
\]  
(A.2)

Remark A.6. (a) The weak derivative is unique, as for the difference \( d = v_i - v'_i \) of two potential candidates, we have \( \int_{\mathbb{R}^k} d \varphi \, d\lambda^k = 0 \) for all \( \varphi \in C^\infty_c(\mathbb{R}^k, \mathbb{R}) \), so by Proposition A.1, \( d \) must be \( 0 \).

(b) Weak derivatives belonging to \( L^2(\lambda^k) \) give rise to the space \( \mathcal{W}^{1,2}(\lambda^k) \) of all functions \( f : \mathbb{R}^k \to \mathbb{R} \) with weak derivatives in \( L^2(\lambda^k) \) of order one endowed with the norm \( \|f\|_{\mathcal{W}^{1,2}(\lambda^k)} := \sum_{k=1}^{\infty} \|\partial_k f\|_{L^2(\lambda^k)}^2 \) which is called Sobolev space of order 2 and 1 for which there is a rich theory.

(c) The following two propositions—under the additional requirement that \( \nabla f \) resp. \( \nabla f \) be in \( L^2(\lambda^k) \), however—may also be found in Maz’ya (1985) Thm.’s 1 and 2).

Proposition A.7. Let \( f \in L^1_m(\lambda^k) \) with a weak gradient \( \nabla f \). Then there is some \( \tilde{f} \), a.c. in \( k \) dimensions with usual gradient \( \nabla \tilde{f} \), such that—up to a \( \lambda^k \)-null set—\( f = \tilde{f} \) and \( \nabla f = \nabla \tilde{f} \).

Proof: Let again \( \Omega_m = [-m, m] \) and consider \( X_m \in \partial_k \) with \( 0 \leq X_m \leq 1 \), \( \phi_m \equiv 0 \) on \( \Omega_{m+1} \) \( \cdot \) \( X_m \equiv 1 \) on \( \Omega_m \), and let \( f_m = f \chi_{X_m} \). Then \( f_m \in L^1(\lambda^k) \) and we have for any \( \phi \in \partial_k \)
\[
-\int f_m \partial_i \phi \, d\lambda^k = -\int f_m \partial_i X_m \phi \, d\lambda^k = \int f \partial_i X_m \phi \, d\lambda^k = \int f \partial_i (X_m \phi) \, d\lambda^k = \int f \partial_i X_m \phi + X_m \partial_i \phi \, d\lambda^k
\]

so that \( f_m \) is weakly differentiable and \( \partial_i f_m = X_m \partial_i f + f \partial_i X_m \in L^1(\lambda^k) \). By Fubini we obtain some \( N_{m,i} \in \mathbb{R}^{k-1} \) with \( \lambda^{k-1}(N_{m,i}) = 0 \) such that \( v_m : \mathbb{R}^{k-1} \to \mathbb{R} \) defined as
\[
v_m(y) := \int_{\mathbb{R}} \tilde{\partial}_i f_m((t;y)) \lambda(dt) \quad \text{for } y \in N_{m,i} \text{ and } 0 \text{ else}
\]
is finite for \( y \in \mathbb{R}^{k-1} \), lies in \( L^1(\lambda^{k-1}) \) and \( \int_{\mathbb{R}^{k-1}} v_m \, d\lambda^{k-1} = \|\tilde{\partial}_i f_m\|_{L^1(\lambda^k)} \). Thus we may define to \( x \in \mathbb{R} \)
\[
F_m((x;y)) := \int_{-\infty}^x \tilde{\partial}_i f_m((t;y)) \lambda(dt) \quad \text{for } y \in N_{m,i} \text{ and } 0 \text{ else}
\]  
(A.3)

Apparently, \( F_m \in L^1_m(\lambda^k) \) and for \( y \in N_{m,i} \), \( x \mapsto F_m((x;y)) \) is a.e. Let \( \phi \in \partial_k \); then Fubini yields
\[
1 := \int_{\mathbb{R}^k} \phi \, d\lambda^k = \int_{N_{m,i}} \tilde{\partial}_i f_m((t;y)) \int_{\mathbb{R}} 1_{[t \leq x]} \phi((x;y)) \lambda(dx) \lambda(dt) \lambda^{k-1}(dy)
\]
Proposition A.8. we see that there is a function \( \tilde{\psi} \) and hence by continuity, for all measure 0. Let \( f \) be a.c. Let \( \phi \) are a.c. Let \( \tilde{\psi} \) is defined as \( \tilde{\psi} = \int_{(x,y) \in \Omega} \phi((x,y)) \lambda(dx) \), which clearly lies in \( \mathcal{D}_k \). Fubini and the definition of weak differentiability entail

\[
I = \int_{\mathbb{R}_+} \int_{\mathbb{R}^k} \tilde{\psi}(t,y) \phi(t,y) \lambda(dt) \lambda^{k-1}(dy) = \int_{\mathbb{R}^k} \phi \tilde{\psi} \, m \, d\lambda^k = - \int_{\mathbb{R}^k} f m \partial_\psi \phi \, d\lambda^k
\]

But, \( \partial_\psi \phi(u) = \partial_\psi \phi(u) \int_{(x,y) \in \Omega} \phi((x,y)) \lambda(dx) = - \psi(u) \phi(u) \), as \( f m \partial_\psi \phi \equiv 0 \), \( f m \equiv f \) we get

\[
I = \int_{\mathbb{R}^k} \phi \, m \, d\lambda^k = \int_{\mathbb{R}^k} \phi \, m \, d\lambda^k.
\]

Because \( \phi \) was arbitrary in \( \mathcal{D}_k \), \( F_m = f_m \mid \mathcal{L}^k \), and by letting \( m \to \infty \) we may extend this to \( \mathbb{R}^k \). Fubini then provides a \( \lambda^k \)-null set \( S_i \), s.t. for \( y \in \mathbb{R}^k \), the projection set \( S_i^{(y)} := \{ x \in \mathbb{R} : (x,y) \in S_i \} \) has \( \lambda \)-measure 0. Let \( N_i := \bigcup_{m} N_{m,i} \), then \( \lambda^k(\bigcup N_i) = 0 \), and for \( y \in \mathbb{N} \), the functions \( x \mapsto F_m((x,y)) \) are a.c., hence continuous in particular. For \( y \in (\bigcup N_i) \), \( x \in (\bigcup S_i^{(y)})^{c} \) and \( k \in \mathbb{N} \), even \( F_m((x,y)) = F_{m+k}((x,y)) \), and hence by continuity, for all \( y \in (\bigcup N_i) \), \( F_m((x,y)) = F_{m+k}((x,y)) \) for all \( x \). Hence, writing again \( u = (t,y) \), this gives a unique function \( \tilde{f}_i \in L_{1,\infty}(\lambda^k) \) defined as

\[
\tilde{f}_i(u) := \begin{cases} 
\lim_{m \to \infty} F_m(u) & \text{for } u \in \mathbb{R}^k, \ y \in (N_i \cup S_i)^c \ \\
0 & \text{else}
\end{cases}
\]

s.t. that \( \tilde{f}_i \) is a.c. w.r.t. \( x_i \) in the sense that there is \( \lambda^{k-1} \)-null set \( \tilde{N}_i \) s.t. for \( y \in \mathbb{N} \), the function \( x \mapsto \tilde{f}_i((x,y)) \) is a.c. By construction, \( \lambda^k(\{ \tilde{f}_i \neq f \} \cup \{ \tilde{f}_j \neq \tilde{f}_i \}) = 0 \). Applying this argument for each \( i = 1, \ldots, k \), we see that there is a function \( f^\prime \) which is a.c. in \( k \) dimensions, s.t. \( \lambda^k(\{ f^\prime \neq f \} \cup \{ \tilde{f}_i \neq f \}) = 0 \). \( \square 

Proposition A.8. Let \( f \in L_{1,\infty}(\lambda^k) \) be a.c. in \( k \) dimensions. If its classical partial derivatives \( \partial_i f \), are extended by 0 on those lines where absolute continuity fails, and the so extended gradient belongs to \( L_{1,\infty}(\lambda^k) \), then there is a weak gradient of \( f \) and the extended gradient can be taken as a version of the weak gradient.

Proof: As \( f \) is a.c. in \( k \) dimensions there exist \( N_i \in \mathbb{B}^{k-1} \) such that for \( y \in N_i \), the functions \( x \mapsto f((x,y)) \) are a.c. Let \( \phi \in \mathcal{D}_k \) and \( y \in N_i \). Then \( x \mapsto \phi((x,y)) \in \mathcal{D}_k \) and thus by integration by parts, for \( y \in N_i \), we have

\[
\int_{\mathbb{R}^k} f((x,y)) \partial_i \phi((x,y)) \lambda(dx) = - \int_{\mathbb{R}^k} \phi((x,y)) \partial_i f((x,y)) \lambda(dx)
\]

Obviously, extending \( f \partial_i \phi \), \( \phi \partial_i f \) by 0 on \( y \in N_i \), these two functions belong to \( L_1(\lambda^k) \). Fubini thus yields a set \( \tilde{N}_i \in \mathbb{B}^{k-1} \), \( \lambda^k(\tilde{N}_i) = 0 \), s.t. for \( y \in \mathbb{N} \), \( x \mapsto f \partial_i \phi((x,y)) \), \( x \mapsto \phi \partial_i f((x,y)) \) belong to \( L_1(\lambda) \). Hence by Fubini \( \int_{\mathbb{R}^k} f \partial_i \phi \, d\lambda^k = - \int_{\mathbb{R}^k} \phi \partial_i f \, d\lambda^k \). As \( \partial_i f \in L_{1,\infty}(\lambda^k) \) by definition of absolute continuity in \( k \) dimensions, this possibly extended \( \partial_i f \) is a weak derivative of \( f \). \( \square 

Remark A.9. Having this “almost” coinciding of weak differentiability and absolute continuity in \( k \) dimensions in mind, we drop the notational difference of weak and classical derivatives.
Appendix B. Proofs

Appendix B.1. Preparations

Before proving Theorem 4.4 some preparations are needed.

We want to parallel the proof given in [Huber (1981)] credited to T. Liggett: The idea is to define for given \( a \in \mathbb{R}^p \) linear functionals \( \bar{T}_{a,i} \) on the dense subset \( \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \) of \( L_2(P_0) \) as

\[
\bar{T}_{a,i} : \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \to \mathbb{R}, \quad \bar{T}_{a,i}(\phi) := \int D_{a,i} \partial_\phi \phi dP_0.
\] (B.1)

Remark B.1. As also true for the one-dimensional location model treated in [Huber (1981)] and in the one-dimensional scale model in [Ruckdeschel and Rieder (2010)], it is not clear a priori whether this is a sound definition, i.e., whether \( \bar{T}_{a,i} \) respect equivalence classes of functions in \( L_2(P_0) \):

As by (Dk) resp. (D1), \( D_{a,i} \) is continuously differentiable, it is bounded on compacts, hence \( D_{a,i} \partial_\phi \phi \in L_1(P_0) \) for any \( \phi \in \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \); but even then, it is still not clear whether (B.1) makes a definition: Take \( x^{(0)} \in \mathbb{R}^k \) so that \( x^{(0)}_i \neq 0 \) and \( P_0 \) Dirac measure for \( \{ x^{(0)} \} \).

Then obviously, \( \phi_0(x) = (x^{(0)})^T(x - x^{(0)}) = 0 \) \( P_0(dx) \)-a.e., but it also holds, \( \partial_\phi \phi_0(x) D_{a,i}(x) = x^{(0)}_i D_{a,i}(x^{(0)}) \) \( P_0(dx) \)-a.e., with the consequence that, although \( \phi_0 = 0 \) \( P_0 \), \( \bar{T}_{a,i}(\phi_0) \neq \bar{T}_{a,i}(0) = 0 \). Of course, \( \phi_0 \) must be modified away from \( x^{(0)} \) to some \( \phi \) s.t. \( \phi \) belongs to \( \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \). Luckily enough, this case cannot occur under condition (i) of Theorem 4.4 as then

\[
\phi = 0 \quad \implies \quad \bar{T}_{a,i}(\phi) = \int \partial_\phi \phi D_{a,i} dP_0 = 0
\] (B.2)

which may be proved just along the lines of the first paragraph of [Ruckdeschel and Rieder (2010) Proof to Thm. 2.2]. Due to linearity of differentiation, evaluated member-wise in an \( P_0 \)-equivalence class, this shows that \( \bar{T} \) respects \( P_0 \)-equivalence classes.

Next we need a lemma showing denseness of certain sets in suitable \( L_2 \)'s. To do so we define for \( i, j = 1, \ldots, k \)

\[
\mathcal{D}_{D_{a,j}} := \{ D_{a,j} \partial_\phi \phi \mid \phi \in \mathcal{D}_k, a \in \mathbb{R}^p \}
\] (B.3)

and recalling that \( K_j := \{ e_j \}D = 0 \) we introduce the decompositions corresponding to (B.4)

\[
P_0 := p_0^{(j)} + \bar{p}_0^{(j)}, \quad \bar{p}_0^{(j)}(\cdot) := P_0(\cdot \cap K_j).
\] (B.4)

Lemma B.2. \( \mathcal{D}_{D_{a,i}} \) is dense in \( L_2(Q) \) for any \( \sigma \)-finite measure on \( \mathbb{R}^k \cap K_j \). In particular it is measure-determining for \( \mathbb{R}^k \cap K_j \).

Proof: Approximating \( f \in L_2(Q) \) in \( L_2(Q) \) by \( f_n := f 1_{\Omega_n} \) with \( \Omega_n = [-n, n]^k \), we may restrict ourselves to \( \Omega_N \) for \( N \) sufficiently large. Thus we have to show that for each interval \( J \subset \mathbb{R}_+ \) and each \( \varepsilon > 0 \), there is a \( \phi \) in \( \mathcal{D}_{D_{a,i}} \) with \( \| \bar{I}_J - \phi \|_{L_2(P_0)} \leq \varepsilon \).

To this end fix \( a \in \mathbb{R}^p \) as \( D_{a,i} \) is continuous, the set \( K_j^* \) is open, hence the countable union of \( k \) dimensional intervals \( J_m := (l(m), r(m)), m \in \mathbb{N} \), with \( l(m) < r(m) \) and \( |D_{a,i}| > 0 \) on \( J_m \).

So it suffices to show that any indicator to an interval \( I = [l, r] \), \( I \subset J_m \) with endpoints s.t. \( Q(\partial I) = 0 \) may be approximated in \( L_2(Q) \) by functions in \( \mathcal{D}_{D_{a,i}} \). But, for given \( \varepsilon > 0 \), Proposition A.1 provides an element \( \phi_0 \in \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \) such that \( \| \phi_0 - \bar{I}_I \|_{Q} < \varepsilon \). By construction its anti-derivative \( \psi_0((x; y_i)) := \int 1_{\{z \in \mathbb{R} \mid \rho_0 \leq z \leq x \}}(z; y_i)D_{a,i}(z; y_i) dz \) lies in \( \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R}) \) hence \( \phi_0 \) in \( \mathcal{D}_{D_{a,i}} \). In particular we may approximate the \( Q \) measure for \( k \)-dimensional intervals disjoint to \( K_j \), which determines \( Q \).
Appendix B.2. Proof of the Main Theorem

(ii) ⇒ (i) of Theorem 4.4

In order to avoid specializing the case $k = 1$, define $V := 0$ there.

Fix $a \in \mathbb{R}^p$, $|a| = 1$. $f$ being a density, $\lambda\{f = 0, \partial_i f \neq 0\} = 0$ for each $i$ and we may write

$$\int (\partial_i \varphi) D_{a;i} dP_0 = \int_{K} (\partial_i \varphi) D_{a;i} dP_0 = \int_{\Gamma(K)} [(\partial_i \varphi) D_{a;i}] \cdot \tau_0 \, dF^{(i)}(\varphi) = \int_{\Gamma(K)} [(\partial_i \varphi) D_{a;i}] \cdot \tau_0 \, f \, d\lambda^k =$$

$$= \int_{K} (\partial_i \varphi) D_{a;i} f_0 |\det \partial_i t_0| d\lambda^k = - \int_{K} \varphi f_0 \partial_i (D_{a;i} |\det \partial_i t_0|) + \varphi D_{a;i} |\det \partial_i t_0| \partial_i f_0 d\lambda^k =$$

$$= - \int_{K^c} \varphi f_0 |\partial_i (\det \partial_i t_0| D_{a;i})| + D_{a;i} \partial_i f_0 |d\lambda^k = - \int \varphi f_0 |\partial_i (\det \partial_i t_0| D_{a;i})| + D_{a;i} \partial_i f_0 |d\lambda^k$$

In equation $(*$) we use that by (ii)(c), on $K^c$, $f$ is a.c. $\lambda^k$ a.e. so that integration by parts integration by parts is available without having to care about border values due to (ii)(b). By (ii)(d) the resulting integrand on the RHS of $*$ is in $L_2(P_0)$. In equation $(**$), we used the fact that in each expression considered above, there appears at least one $D_{a;i}$ or a derivative $\partial_i D_{a;i}$; Lemma A.4 applies and hence

$$\lambda^k \{ (D_{a;i} = 0), \{\partial_i D_{a;i} \neq 0\}\} = 0$$

Representations (4.3) and (4.4):

Writing out $\partial_i f_0 = (\partial_i t_0) (\partial_i f) \circ t_0$, we see that

$$D^2_{a;i} \partial_i f_0 = a^T (\partial_i t_0) f (\partial_i t_0) (\partial_i f) \circ t_0 = a^T (\partial_i t_0) (\partial_i f) \circ t_0 = a^T \partial_0 f_0$$

(B.5)

Thus we get

$$\sum_{i=0}^k \frac{\partial_i |D_{a;i} |}{|\det \partial_i t_0|} + D^2_{a;i} \partial_i f_0 \frac{\partial_0}{f_0} = \sum_{i=0}^k \frac{a^T \partial_0 f_0}{|\det \partial_i t_0|} = \sum_{i=0}^k a^T \partial_0 f_0 = a^T \Lambda_0,$$

so $\Lambda_0 \in L_2^2(P_0)$ by (ii)(d) and hence

$$\left( \int \varphi^2 dP_0 + \varphi V a dP_0 \right)^2 \leq \left( \int \varphi a^2 \partial_0 f_0 P_0 \right)^2 \leq \int |a^T \Lambda_0|^2 dP_0 \int \varphi^2 dP_0,$$

which shows that $\mathcal{F}(F; a) \leq \int (a^T \Lambda_0)^2 dP_0$. The upper bound may be approximated by a sequence $\varphi_n \in \mathcal{D}_k$ tending to $a^T \Lambda_0$ in $L_2^2(P_0)$ entailing (4.5) and (4.6).

(i) ⇒ (ii) in Theorem 4.4

We will give a proof largely paralleling [Huber (1981)], although we may skip some of his arguments.

**Well defined operators and Riesz-Fréchet:** We consider the linear functionals $\tilde{T}_{a;i}$ from (B.1), defined on the dense subset $\mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R})$ of $L_2(P_0)$, which are well defined due to (B.2). In particular $\tilde{T}_{a;i}$ are bounded linear operators with squared operator norms bounded by $\mathcal{F}_0(F; a)$, hence can be extended by continuity to continuous linear operators $T_{a;i} : L_2(P_0) \to \mathbb{R}$ with the same operator norms. Thus Riesz Fréchet applies, yielding generating elements $g_{a;i} \in L_2(P_0)$ s.t.

$$T_{a;i}(\varphi) = - \int g_{a;i} \varphi dP_0 \quad \forall \varphi \in L_2(P_0) \quad \text{and} \quad \|g_{a;i}\|^2_{L_2(P_0)} = \|T_{a;i}\|| (B.6)$$

We conclude inductively for $i = 1, \ldots, k$.

**Using Fubini:** We have for $\varphi \in \mathcal{D}_k$

$$T_{a;1}(\varphi) = \int D_{a;1} \partial_1 \varphi dP_0$$

(B.7)
On the other hand by the fundamental theorem of calculus,

\[ T_{a;1}(\varphi) = -\int \varphi g_{a;1} \, dP_\theta = \int_{\mathbb{R}^4} \chi_{\{z_1 \geq y_1\}} \, \partial_{z_1} \varphi(x_1, y_2, k) \lambda(dx_1) g_{a;1}(y) P_\theta(dy) \]

Now for each compact \( A \), the integrand \( h_1(x_1, y; A) = \chi_{\{z_1 \geq y_1\}} \, g_{a;1}(y) \) is in \( L_1(\lambda(dx_1) \otimes P_\theta(dy)) \). Fubini for Markov kernels thus yields a \( \lambda \otimes P_{\theta;2;k} \)-null set \( N_1 \) such that for \((x_1, y_2, k) \in N_1^c \), \( x_1 \in A \), the function \( h_1(y_1) := h_1(x_1, y; A) \) belongs to \( L_1(P_{\theta;1|2;k}(dy|y_2, k)) \). We now define for \( D_{a;1} \neq 0 \) the function \( p^{(a;1)}_{1|2;k} \) as

\[
[D_{a;1} p^{(a;1)}_{1|2;k}](x_1, y_2, k) := \begin{cases} 
\int h_1(y_1) P_{\theta;1|2;k}(dy_1|y_2, k) & \text{for } (x_1, y_2, k) \in N_1^c, x_1 \in A \\
0 & \text{else}
\end{cases} \tag{B.8}
\]

where obviously the dependence in \( A \) is such that for another compact \( A' \supset A \), \( p^{(a;1)}_{1|2;k} = p^{(a;1)}_{1|2;k} \chi_{A}(x_1) \). Hence for arbitrary \( x_1 \) take \( A \) such that \( x_1 \in A \) and eliminate the index \( A \) in the superscript where it is clear from the context. We also note that by Cauchy-Schwarz, for \( p_{\theta;2;k}(dy_2, k) \)-a.e. \( y_2, k \),

\[
||D_{a;1} p^{(a;1)}_{1|2;k}(x_1, y_2, k)||^2 \leq \int g_{a;1}(y)^2 P_{\theta;1|2;k}(dy_1|y_2, k) < \infty \tag{B.9}
\]

**Getting rid of the dependence on \( a \):** To understand, how \( p^{(a;1)}_{1|2;k} \) is related to \( p^{(a';1)}_{1|2;k} \) for \( a \neq a' \in \mathbb{R}^p \), we consider again \( \text{[B.5], [B.8]} \): Both sides of the latter must be of form \( \tilde{W} \cdot a \) for some \( \mathbb{R}^p \) valued \( \tilde{W} \) independent of \( a \); in particular

\[
g_{a;1} = w_1 a \tag{B.10}
\]

for some \( w_1 \in L^2_{\mathbb{F}}(P_{\theta}) \). Hence

\[
p^{(a;1)}_{1|2;k} = p^{(a';1)}_{1|2;k} \quad \text{on } \{D_{a;1}^{(1)} \neq 0\} \cap \{D_{a';1}^{(2)} \neq 0\} \tag{B.11}
\]

and, as we only need \( p \) orthogonal values of \( a \) to specify \( w_1 \), we arrive at a maximally extended \( p^{(1)}_{1|2;k} \) defined on \( K_1^c = \{D_{a} \neq 0\} \). Also, for \( P_{\theta;2;k}(dy_2, k) \)-a.e. \( y_2, k \),

\[
||D_{a;1} p^{(1)}_{1|2;k}(x_1, y_2, k)||^2 \leq |a|^2 \int |w_1(y)|^2 P_{\theta;1|2;k}(dy_1|y_2, k) \tag{B.12}
\]

\( p^{(1)}_{1|2;k} \) is a density: Plugging in this maximal definition, we get for \( \varphi \in D_k \), using \( A = \text{supp}(\varphi) \),

\[
T_{a;1}(\varphi) = \int D_{a;1} \partial_{z_1} \varphi \, dP_\theta = \int [D_{a;1} \partial_{z_1} \varphi p^{(1)}_{1|2;k}(x_1, y_2, k)] \lambda(dx_1) P_{\theta;2;k}(dy_2, k) \]. \tag{B.13}

where integrability of the integrands follows from Remark 2.3(c) and (C1)(Ck), and for the right one from \([B.12] \), which also entails that \( P_{\theta;2;k}(dy_2, k) \)-a.s., \( x_1 \mapsto D_{a;1} p^{(1)}_{1|2;k} \) is the \( \lambda \)-density of a \( \sigma \)-finite signed measure. Hence, we have shown that \( P_{\theta}(dx_1, dy_2, k) \) and \( p^{(1)}_{1|2;k}(x_1, y_2, k) \lambda(dx_1) P_{\theta;2;k}(dy_2, k) \) when restricted to \( K_1^c \) define the same functional on the set \( D_{D;1,1,k} \), which is measure-determining for \( B^k \cap K_1^c \) due to Lemma 2.3.

Therefore, the restriction to compacts \( A \) can be dropped entirely, and we may work with \( A = \mathbb{R} \). Using Fubini once again, we see that on \( K_1^c \), there is a \( P_{\theta;2;k}(dy_2, k) \)-null set \( N_1 \), s.t. for fixed \( y_2, k \in N_1^c \), the function \( p^{(1)}_{1|2;k}(x_1, y_2, k) \) is a Lebesgue density of the regular conditional distribution \( P^{(0)}_{\theta;1|2;k}(dx_1|y_2, k) \), hence non negative and in \( L_1(\lambda) \).

**Replacing \( K_1 \) by \( K \):** Similarly as for the dependence on \( a \), we may extend the definition of \( p^{(1)}_{1|2;k}(x_1, y_2, k) \) to the set \( K^c \): Any \( \partial_{\phi} \varphi \) for \( \varphi \in D_k \) may also be interpreted as \( \partial_{\phi} \varphi \) for some \( \phi \in D_k \). More specifically,
Assume we have already shown that there is a \( \phi \) the section-wise defined function \( i \) so we may indeed speak of a maximally extended set—for simplicity again \( \tilde{\phi} \). Thus using the induction assumptions we proceed as before, i.e.; define \( \tilde{B} K \) density defined on \( a \) and outside \( a \). Now \( \tilde{p} \) being Lebesgue densities of \( \tilde{p}^0 \), there is a \( P_{\theta,2:K} \)-null set—for simplicity again \( \tilde{N}_1 \)—such that for \( y \in \tilde{N}_1 \), for each pair \( j_1 \neq j_2 \),

\[
p_{ij}(x,y)_1 = \tilde{p}_{ij}(x,y)_1 \quad [\lambda(dx)] \quad \text{on} \ K'_i \cap K'_j, \tag{B.14}
\]

so we may indeed speak of a maximally extended \( p_{1:2:k} \) defined on \( K' \).

Assume we have already shown that there is a \( P_{\theta,1:k} \)-null set \( \tilde{N}_1 \) such that for \( y_{1:k} \in \tilde{N}_1 \), \( P_{\theta}^0 \) admits some conditional density,

\[
p_{1:j-1:k}(x_{1:j-1},y_{1:k})\lambda^{-1}(d\gamma_{j-1}) = p_{\theta}^0(n_{j-1-1:k})(d\gamma_{j-1-1},y_{1:k}). \tag{B.15}
\]

Arguing just as for \( j = 1 \), we get

\[
T_{\theta}(\phi) = \int D_{\theta} \phi \ d\theta P_{\theta} = \int \mathbb{I}_{\{D_{\theta} \neq 0\}} D_{\theta} \phi \ d\theta P_{\theta} = - \int \phi D_{\theta} \ d\theta P_{\theta}.
\]

Thus using the induction assumptions we proceed as before, i.e.; define \( \tilde{h}(x,y;A) \) for some compact \( A \), a the section-wise defined function \( h_i(y) := \tilde{h}(x,y;A) \), the function \( p_{\theta}^{(a)} \) which extends to \( \tilde{R} \), giving \( p_{\theta}^{(a)} \), and where the dependence on \( a \) may be dropped, giving \( p_{\theta}^{(a)} \). As this defines the same functional on the set \( D_{\theta} \) as the Markov kernel \( p_{\theta}^{(a)} \) by Lemma B.2, \( p_{\theta}^{(a)} \) is a conditional density defined on \( K' \). Using the coordinate permutation argument to drop the dependence on \( K_i \), we obtain \( p_{\theta}^{(a)} \), defined on \( K' \).

Hence the induction is complete, and we have shown that \( P_{\theta}^0 \) admits a \( \lambda \) for any fixed \( a \in \mathbb{R}^p \) with \( p_{\theta}^{(a)} \) and \( p_{\theta}^{(a)} \), we see that by Fubini, for \( y \) outside a \( P_{\theta,1:k} \)-null set,

\[
|D_{\theta} P_{\theta,1:k} ((z:y))_i | := \int_{-\infty}^{\infty} g_{\theta}((z:y))_i P_{\theta,1:k} (dz) + \gamma_{\theta} (z:y))_i \tag{B.16}
\]

with

\[
\gamma_{\theta} (z:y))_i := \int_{-\infty}^{\infty} g_{\theta}((z:y))_i P_{\theta,1:k} (dz) \tag{B.17}
\]

We next show that for fixed \( a \in \mathbb{R}^p \) and fixed \( y \) outside a \( P_{\theta,1:k} \)-null set, the value of \( \gamma_{\theta} \equiv 0 \):

To this end we show that for any Borel subset \( B \) of \( K \) or equivalently for any proper or improper interval \( I = [-r, r] \subset K \),

\[
\int g_{\theta}((z;y))_i \tilde{P}_{\theta,1:k} (dz) = 0
\]

Of course, \( \int d\tilde{P}_{\theta,1:k} (dz) = 0 \), \( [P_{\theta,1:k}] \). Consider \( \phi \) with \( 0 \leq \phi \leq 1 \), \( \phi \equiv 1 \) on \( I \) and \( \phi \equiv 0 \) for \( \{x \in \mathbb{R}^d | \text{dist}(x,I) > 1/n \} \), and \( |\partial_x \phi| \leq 6n \). The last bound is chosen according to the bound \( |\phi| \leq 2c_{\theta} \delta \) from Proposition A.1. Then \( \phi_n \to \phi \) pointwise, hence by dominated convergence and Cauchy-Schwartz we get

\[
\left\| \phi_{n}^{2} d\tilde{P}_{\theta,1:k} \right\|_{L^1} = o(n^0), \quad \left\| g_{\theta} \phi_{n} d\tilde{P}_{\theta,1:k} \right\|_{L^1} = o(n^0).
\]
On the other hand let \( C := \max \{ |\partial_x D_{a;i}| \mid x \in \text{supp}(\phi_1) \} \). Then as \( D_{a;i} = 0 \) on \( I \), and because \( \text{supp}(\partial_x \phi_a) \subset \{ x \in \mathbb{R}^k \mid 0 < \text{dist}(x,I) < 1/n \} \), we have for \( x \in \text{supp}(\phi_1) \) that \( |D_{a;i}(x)| \leq C|x| \leq C/n \), and hence

\[
|\partial_x \phi_a D_{a;i}| \leq 6C_1(\text{supp}(\phi_1) \cap F).
\]

Thus, due to the shrinking of \( \{ \text{supp}(\phi_a) \cap F \} \), for \( x \not\in I, \ |\partial_x \phi_a D_{a;i}(x)| \to 0.\) Furthermore \( |\partial_x \phi_a D_{a;i}|(x) = 0 \) on \( I \), as \( D_{a;i}(x) = 0 \) on \( I \subset K \) by definition, hence also \( |\partial_x \phi_a D_{a;i}| \to 0 \) pointwise and with dominated convergence

\[
\int D_{a;i} \partial_x \phi_a dP_{\theta|i} = o(n^0).
\]

So we have

\[
o(n^0) + \int g_{a;i}(z) F_{\theta|i}^{(0)}(dz) = \int g_{a;i} \partial_z \phi_a dF_{\theta|i}^{(0)} = -\int \partial_z \phi_a D_{a;i} dP_{\theta|i} - \int g_{a;i} \partial_z \phi_a dF_{\theta|i}^{(0)} = o(n^0),
\]

which implies \( \delta_{a;i} \equiv 0 \) and hence, integrating by \( F_{\theta|i}^{(0)} \) over any \( A \in \mathbb{B}^{k-1}, \ g_{a;i} = 0 \) \( F_{\theta|i}^{(0)} \). Similarly, we obtain

\[
g_{a;i,j} = \int D_{a;i} \partial_j \phi_a dP_{\theta|i} = \int \partial_j \phi_a D_{a;i} d\lambda^k = \int \phi g_{a;j,i} dP_{\theta|i} - \int \phi g_{a;i,j} d\lambda^k.
\]

This also entails that

\[
\int \partial_i \phi D_{a;i} dP_{\theta} = \int \partial_i \phi D_{a;i} d\lambda^k = -\int \phi g_{a;i,j} dP_{\theta|i} - \int \phi g_{a;j,i} d\lambda^k.
\]

**Application of Proposition A.7** From \( B.19 \), we get \( \int \partial_i \phi D_{a;i} d\lambda^k = -\int \phi g_{a;j,i} dP_{\theta|i} \) for all \( \phi \in \mathcal{D}_k \). By Definition \( A.5 \), \( g_{a;i,j} dP_{\theta|i} \) thus is the weak derivative of \( D_{a;i} d\lambda^k \) w.r.t. \( x_j \). By Proposition \( A.7 \), there is a modification of \( D_{a;i} d\lambda^k \) on a \( \lambda^k \)-null set such that this modification—for simplicity again denoted by \( D_{a;i} d\lambda^k \) in \( k \) dimensions. Hence, for \( \lambda^k \) a.e. \( x, \ D_{a;i} d\lambda^k \) is differentiable w.r.t. \( x_j \) in the classical sense with a derivative coinciding with \( g_{a;i,j} d\lambda^k \) up to a \( \lambda^k \)-null set.

As \( D_{a;i} \) is continuously differentiable, \( \lambda^k \) is differentiable on \( K^c \) for \( \lambda^k \) a.e. \( x \), and using again all the different \( D_{a;i}, j = 1, \ldots, k \), the same is even true on \( K^c \).

**Proof of (ii)(a)–(d):** Defining for \( \theta \in \Theta \)

\[
f(\theta) := (p_\theta / |\det \partial_x t_\theta|) \circ t_\theta,
\]

and recalling that \( P_\theta = F \circ t_\theta \), we see that by the Lebesgue transformation formula, \( f(\theta) \) must be a density of \( F \), hence the index \( \theta \) may be dropped, and (ii)(a) follows. Once again by the transformation formula,

\[
p_\theta = |\det \partial_x t_\theta| (F \circ t_\theta) = |\det \partial_x t_\theta| F_\theta.
\]

and thus (ii)(c) holds. For (ii)(b) we consider \( \kappa \) defined analogously as for \( k = 1 \) as inverse to \( \ell \) from \( \$7,3 \). We lift \( \mathcal{D}_k \) to \( [0,1]^k \), giving

\[
\left( \int_{[0,1]^k} \psi \kappa q_\theta d\lambda \right)^2 \leq \mathcal{F}_\theta(F) \int_{[0,1]^k} \psi^2 d(\ell(P_\theta)) \quad \forall \psi \in C_c((0,1)^k, \mathbb{R}),
\]

where \( q_\theta = p_\theta \circ \kappa \), and we have to show that \( |\kappa q_\theta| u = 0 \) for \( u \in \partial([0,1]^k) \). We only show \( u_1 = 1 \), all other cases follow similarly. Let \( \psi \in \mathcal{D}_k, \ \psi_\theta \to I_{[1]} \times [0,1]^{k-1} \) in \( L_2(\ell(P_\theta)) \) and pointwise. Then by Fubini and by integration by parts

\[
\int_{[0,1]^{k-1}} \int_0^1 \left[ \partial_1 (\psi_\theta) \kappa q_\theta \right] (\psi_\theta) dx \lambda (dx) \lambda^{k-1} (dy) =
\]

\[
= \int_{[0,1]^{k-1}} \left[ \psi_\theta \kappa q_\theta \right]_0^1 - \int_0^1 \psi \kappa \psi_\theta [\ell(P_\theta)]_{1;2,k} (ax,y) [\ell(P_\theta)]_{2;2,k} (dy)
\]

20
But \( \int_{[0,1]} \psi^2_a d[\ell(P_\theta)] \to 0 \) entails by Fubini. \( \int_{[0,1]} \psi^2_a d[\ell(P_\theta)] \mid_{1+2} \to 0 \) \( [\ell(P_\theta)]_{1+2}(dy) \) a.e. and by Cauchy-Schwartz that also \( \int_{[0,1]} g^0 \psi_\theta d[\ell(P_\theta)] \mid_{1+2} \to 0 \) and hence
\[
\left( \psi_\theta \kappa_\theta \right)(1:y_1) + o(n^0))^2 \leq o(n^0) \quad [\ell(P_\theta)]_{1+2}(dy)
\]
and due to continuity of \( \psi_\theta \kappa_\theta \), (ii)(b) follows. For (ii)(d) we proceed as in part (ii) \( \Rightarrow \) (i)
\[
\left( \sum e_{i,j} - V_\theta \right) p_\theta = \left( \sum \partial_i [D_{ax} p_\theta] \right) - V_\theta p_\theta = \theta_p \frac{\partial_i [D_{ax} \partial_j t_\theta]}{f_\theta} - V_\theta p_\theta = \theta_p \left( \frac{\partial_i [D_{ax} \partial_j t_\theta]}{f_\theta} + \frac{\partial_i [D_{ax} \partial_j t_\theta]}{f_\theta} \right) (B.22)
\]
Now (ii)(c) follows from (B.22) and the fact that \( V_\theta \) and all \( g_{ax} \) are in \( L_2(P_\theta) \), and assertions (4.5) and (4.6) from (B.23).

The next corollary shows that \( K \) is uninformative for our problem in the sense that \( \tilde{\theta}_\theta^{0} \) a.e. \( \Lambda_\theta = 0 \).

**Corollary B.3.** Under the assumptions of Theorem 4.4 setting
\[
\Lambda_\theta := - V + \sum w_i \quad (B.24)
\]
with \( w_i \) from (B.10) (for \( i = 1 \)) and respectively defined otherwise, it holds that
\[
\Lambda_\theta = 0 \quad (B.25)
\]

**Proof:** (B.24) is defined according to (B.23) on \( K^c \), and as (B.18) entails \( \sum w_i = 0 \) \( [\tilde{\theta}_\theta^{0}] \), the assertion is a direct consequence of \( x \in K \iff D(x) = 0 \iff \partial_\theta t_\theta(x) = 0 \quad \text{and} \quad V(x) = 0 \). \( \square \)

**Appendix B.3. Proofs of Sections 6**

For the proof of Proposition 6.1 we need two lemmas:

**Lemma B.4.** The multivariate location model \( \text{(3.2)} \) is \( L_2 \) -differentiable if it is “partially” in each coordinate separately, i.e.;
\[
\int \left( \sqrt{\lambda}(x_1, \ldots, x_j + h, \ldots, x_k) - \sqrt{\lambda}(x)(1 - \frac{1}{2} \Lambda_{j, i}(x)) \right)^2 \lambda^k(x) = 0 \quad (B.26)
\]

**Proof to Lemma B.4** Garel and Hallin (1995) Lemma 2.1. \( \square \)

**Lemma B.5.** The multivariate location model \( \text{(3.6)} \) is \( L_2 \) -differentiable if it is “partially” in each coordinate separately, i.e.; for each \( i, j = 1, \ldots, k \) and each \( A = A^T \in \mathbb{R}^{k \times k} \)
\[
\int \left( \sqrt{\text{det}(I_k + h \delta_{i,j} A)} \sqrt{\lambda}(I_k + h \delta_{i,j} A)(x) - \sqrt{\lambda}(x)(1 + \frac{1}{2} \Lambda_{i, j}(x)) \right)^2 \lambda^k(x) = 0 \quad (B.27)
\]
where \( \delta_{i,j} \) is the matrix in \( \mathbb{R}^{k \times k} \) with but 0 entries except at position \( i, j \).

**Proof to Lemma B.5** With obvious translation we may parallel Garel and Hallin (1995) Lemma 2.1. A proof is given in Ruckdeschel (2001) Lemma B.3.3\( . \) \( \square \)

**Proof to Proposition 6.1** Putting together Lemmas B.4 and B.5 we have reduced the problem to the respective questions in the one dimensional location resp. scale model, which is proven in Häjek (1972) (one-dimensional location) and Swensen (1980) Ch.2, Sec.3) (one-dimensional scale); Ruckdeschel and Rieder (2010) Prop. 3.1) in addition shows that in the pure scale case, we may allow for mass in 0. \( \square \)
Appendix B.4. Proofs of Section 7

Proof to Proposition 7.4: For fixed $0 \neq a \in \mathbb{R}^p$, the proof goes through word by word as in [Huber (1981)], simply replacing $f_j'$ by $a^T \partial_\theta f_j$ and $f_1$ by $f_1'$: By a monotone convergence argument it is shown that we may differentiate twice under the integral sign, giving

$$\frac{d^2}{dt^2} \mathcal{I}_\theta(F; a) = \int \left( \frac{a^T \partial_\theta f_1}{f_1} \right)^2 \partial_\theta^2 F(x) \, dx.$$  

So we conclude that $a^T \partial_\theta \log f_0 = a^T \partial_\theta \log f_1 \lambda^k (dx)$ a.e., i.e.,

$$a^T \partial_\theta \log f_0 \frac{\nabla F}{F} \circ t_\theta(x) + a^T \partial_\theta \log |\det \partial_\theta t_\theta(x)| = a^T \partial_\theta \log f_1 \frac{\nabla F}{F} \circ t_\theta(x) + a^T \partial_\theta \log |\det \partial_\theta t_\theta(x)|,$$

where due to (d) up to a $\lambda^k$-null set $\int \frac{\nabla F}{F} \circ t_\theta = \int \frac{\nabla f}{f} \circ t_\theta$, and hence up to a $\lambda^k$-null set $\int \log f_0 = \int \log f_1$.

Integrating this out w.r.t $x_i$, we get by (c) that $f_0(x) = c_{i(x)} f_i(x)$ for $\lambda^{k-1}$ almost all $x_i$. Varying $i$, we see that for some $c > 0$, $c_{i(x)} = c$ for all $i = 1, \ldots, k$ and for $\lambda^k$ almost all $x$, and hence

$$\mathcal{I}_\theta(F; a) = \int \left( \frac{a^T \partial_\theta f_1}{f_1} \right)^2 f_1 \, d\lambda^k = \int \left( \frac{a^T \partial_\theta f_0}{f_0} \right)^2 f_0 \, d\lambda^k = c \mathcal{I}_\theta(F_0; a)$$

and $c = 1$. As this holds for any $0 \neq a \in \mathbb{R}^p$, the assertion for $\mathcal{I}_\theta(F)$ follows. \qed

Proof to Proposition 7.3: As by Proposition 7.4 for any $a \in \mathbb{R}^p$ the mapping $F \mapsto \mathcal{I}_\theta(F; a)$ is weakly lower-semicontinuous, the same goes for the following, recursively defined mappings: Let $a_1 \in \mathbb{R}^p$, $|a_1| = 1$ realize

$$\hat{\mathcal{I}}_{\theta, 1}(F) := \mathcal{I}_\theta(F) = \max_{a \in \mathbb{R}^p, |a| = 1} \mathcal{I}_\theta(F; a),$$

and for $i = 2, \ldots, k$, assuming $a_j$ already defined for $j = 1, \ldots, i-1$, let $a_i \in \mathbb{R}^p$, $|a_i| = 1$ realize

$$\hat{\mathcal{I}}_{\theta, i}(F) := \max_{a \in \mathbb{R}^p, |a| = 1, a \perp \{a_j\}_{j<i}} \mathcal{I}_\theta(F; a).$$

Then each of the $\hat{\mathcal{I}}_{\theta, i}(F)$, $i = 1, \ldots, k$ is weakly lower-semicontinuous by the same argument as $\mathcal{I}_\theta(F)$ and is strictly positive by assumption. Hence for each $i = 1, \ldots, k$, the mapping $F \mapsto 1/\hat{\mathcal{I}}_{\theta, i}(F)$ is weakly upper-semicontinuous, and so is the sum $\sum_1/\hat{\mathcal{I}}_{\theta, i}(F)$. But this sum is just the trace of $[\mathcal{I}_\theta(F)]^{-1}$. The corresponding statement as to the attainment of the maximum is shown just as Corollary 7.3. \qed

Acknowledgement

The author wants to thank Helmut Rieder for his helpful suggestions.

References


