Probability Theory and Related Fields **141** (2008) 213–246

Generalized maximum likelihood estimates for exponential families*

Imre Csiszár† *and František Matúš*‡

This paper is dedicated to the memory of Albert Perez (1920-2003).

Abstract. For a standard full exponential family on $\mathbb{R}^d$, or its canonically convex subfamily, the generalized maximum likelihood estimator is an extension of the mapping that assigns to the mean $a \in \mathbb{R}^d$ of a sample for which a maximizer ϑ^* of a corresponding likelihood function exists, the member of the family parameterized by ϑ^*. This extension assigns to each $a \in \mathbb{R}^d$ with the likelihood function bounded above, a member of the closure of the family in variation distance. Its detailed description, complete characterization of domain and range, and additional results are presented, not imposing any regularity assumptions. In addition to basic convex analysis tools, the authors' prior results on convex cores of measures and closures of exponential families are used.

1. Introduction

1.1. Let μ be a nonzero Borel measure on $\mathbb{R}^d$ whose log-Laplace transform

$$\Lambda(\vartheta) = \Lambda_\mu(\vartheta) \triangleq \ln \int_{\mathbb{R}^d} e^{\langle \vartheta, x\rangle} \, \mu(dx) \,, \qquad \vartheta \in \mathbb{R}^d \,,$$

has the effective domain $\mathsf{dom}(\Lambda) = \{\vartheta \colon \Lambda(\vartheta) < +\infty\}$ nonempty. The *full exponential family* $\mathcal{E} = \mathcal{E}_\mu$, determined by μ and the identity mapping as canonical statistic, consists of the probability measures (pm's) $Q_\vartheta = Q_{\mu,\vartheta}$, $\vartheta \in \mathsf{dom}(\Lambda)$, with μ-densities $x \mapsto e^{\langle \vartheta, x\rangle - \Lambda(\vartheta)}$. Here, ϑ is called canonical parameter. The mapping $\vartheta \mapsto Q_\vartheta$ may be many-to-one.

Throughout this paper the symbol Ξ denotes a nonempty convex subset of $\mathsf{dom}(\Lambda)$ and $\mathcal{E}_\Xi = \mathcal{E}_{\mu,\Xi}$ denotes the *canonically convex exponential family* $\{Q_\vartheta \colon \vartheta \in \Xi\}$. The full family $\mathcal{E}$ is regarded as the family $\mathcal{E}_\Xi$ with $\Xi = \mathsf{dom}(\Lambda)$, even if $\mathcal{E} = \mathcal{E}_\Xi$ may hold also for other choices of Ξ. All results in this paper are stated for general Ξ, full exponential

*This work was supported by the Hungarian National Foundation for Scientific Research under Grant T046376 and by Grant Agency of Academy of Sciences of the Czech Republic under Grant IAA 100750603.

AMS 2000 Mathematics Subject Classification. Primary 60A10; secondary 62H12, 62B10.

Key words and phrases. Accessible face, convex core, exponential family, Kullback-Leibler information divergence, log-convexity, maximum likelihood, partial mean, variation distance closure.

†I. Csiszár is with A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, H-1364 Budapest, P.O.Box 127, Hungary (`csiszar@renyi.hu`).

‡F. Matúš is with Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic (`matus@utia.cas.cz`).

families are covered as the instance $\Xi = \mathit{dom}(\Lambda)$. Exponential families with a canonical statistic different from the identity mapping will not be addressed, but the results easily extend to those, via reduction by sufficiency to "standard" families as above.

For $a \in \mathbb{R}^d$ a maximizer ϑ^* of the function $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$ subject to $\vartheta \in \Xi$ has a statistical interpretation when a equals the mean of an i.i.d. sample from a probability measure Q_ϑ with $\vartheta \in \Xi$ unknown. Then ϑ^* is a maximum likelihood estimate (MLE) of the unknown parameter, from this sample. As well known [1], a sufficient condition for the existence of MLE is the equality of a to the mean of Q_θ for some $\theta \in \Xi$. Then $\vartheta^* = \theta$ maximizes the above function even when ϑ ranges over the whole $\mathit{dom}(\Lambda)$, and

$$\big[\langle \vartheta^*, a \rangle - \Lambda(\vartheta^*)\big] - \big[\langle \vartheta, a \rangle - \Lambda(\vartheta)\big] = D(Q_{\vartheta^*} \| Q_\vartheta)\,, \qquad \vartheta \in \Xi\,.$$

Here, D denotes Kullback-Leibler information divergence (I-divergence or relative entropy), defined for any pm's P and Q by

$$D(P\|Q) \triangleq \begin{cases} \int \ln \frac{dP}{dQ}\ dP\ , & \text{if } P \ll Q, \\ +\infty\ , & \text{otherwise.} \end{cases}$$

The authors have recently shown [6, 5] that for any $a \in \mathbb{R}^d$ such that

(1) $$\Psi^*(a) = \Psi^*_{\mu,\Xi}(a) \triangleq \sup_{\vartheta \in \Xi} \big[\, \langle \vartheta, a \rangle - \Lambda(\vartheta)\, \big]$$

is finite, thus $a \in \mathit{dom}(\Psi^*)$, there exists a unique pm $R^*(a)$ with the property

(2) $$\Psi^*(a) - \big[\langle \vartheta, a \rangle - \Lambda(\vartheta)\big] \geqslant D(R^*(a)\|Q_\vartheta)\,, \qquad \vartheta \in \Xi\,.$$

This $R^*(a) = R^*_{\mu,\Xi}(a)$ has been called *generalized maximum likelihood estimate* (GMLE). In the case of existence of an MLE ϑ^*, a direct substitution gives that $R^*(a) = Q_{\vartheta^*}$, thus ϑ^* parameterizes the GMLE. The inequality (2) implies that for each $\vartheta \in \Xi$ satisfying $\langle \vartheta, a \rangle - \Lambda(\vartheta) \geqslant \Psi^*(a) - \varepsilon$ (such ϑ may be called an ε-MLE), the pm Q_ϑ belongs to an information divergence ball of radius ε centered at $R^*(a)$. The concept of GMLE should be of interest in those cases when no MLE exists (not even in the slightly modified sense discussed in Section 3, corresponding to the closure of a projection of the set Ξ), but the GMLE exists, thus it can serve as a substitute of MLE. In these cases, the GMLE does not belong to the exponential family $\mathcal{E}$, see the passage after Theorem 3.2, although it always belongs to $\mathit{cl}_v(\mathcal{E}_\Xi)$, the closure of $\mathcal{E}_\Xi$ in variation distance, see Remark 1.2 below. This means that for GMLE to go beyond MLE it is necessary that $\mathcal{E}$ be not closed in variation distance, which excludes statistical models of continuous type. For models of discrete or mixed type, however, it is not uncommon that the log-likelihood function is bounded but has no maximizer; then GMLE provides a remedy to the nonexistence of MLE.

Remark 1.1. The notation points to the fact that Ψ^* is the convex conjugate of the function Ψ equal to Λ on Ξ and to $+\infty$, otherwise. Thus, (2) sharpens the Fenchel inequality for Ψ, stating that the left-hand side of (2) is nonnegative. Note that if a full exponential family is considered then $\Psi = \Lambda$, $\Psi^* = \Lambda^*$.

Remark 1.2. For any sequence ϑ_n in Ξ with $\langle \vartheta_n, a \rangle - \Lambda(\vartheta_n) \to \Psi^*(a) < \infty$ it follows from (2) that the sequence Q_{ϑ_n} converges to $R^*(a)$ in reversed information (rI-) divergence, $D(R^*(a)\|Q_{\vartheta_n}) \to 0$, and hence by the Pinsker inequality also in variation distance. This shows that the GMLE $R^*(a)$ is uniquely determined by (2), and belongs to $\mathit{cl}_v(\mathcal{E}_\Xi)$. This closure has been recently described in detail [8].

Remark 1.3. The supremum in (1) can be finite also if a is outside the convex support of μ, denoted by $\mathit{cs}(\mu)$, though not for a full family since $\mathit{dom}(\Lambda^*) \subseteq \mathit{cs}(\mu)$ [1, Theorem 9.1]. Then the maximization has no direct statistical interpretation since the mean of a sample from a distribution in the family $\mathcal{E}_\Xi$ belongs to $\mathit{cs}(\mu)$ with probability 1. As explained below, it is still useful to consider MLE and GMLE also when $a \notin \mathit{cs}(\mu)$, and even when a is outside the affine hull of $\mathit{cs}(\mu)$, denoted by $\mathit{aff}(\mu)$. In that case, different parameter sets Ξ can give rise to the same family of pm's $\mathcal{E}_\Xi$ but to different GMLE's, see Example 3.5 in Section 3. For this reason, the choice of Ξ is left free in this paper, unlike in [8] where from the possible parameter sets for the family $\mathcal{E}_\Xi$ a natural one is selected.

1.2. The goal of this paper is to study properties of the GMLE in full generality. This generality is not for its own sake. Indeed, suppose a characterization of $R^*(a)$ were required only for full families $\mathcal{E}_\mu$ in minimal representation, that is, such that both $\mathit{cs}(\mu)$ and $\mathit{dom}(\Lambda_\mu)$ have nonempty interiors. As illustrated in Example 1.4 below, even in that case there may exist a set A that contains the sample mean with positive probability (for all sample sizes) such that the characterization of $R^*(a)$ for $a \in A$ necessitates the maximization of $\langle\vartheta, a\rangle - \Lambda_\nu(\vartheta)$ subject to $\vartheta \in \Xi$ where ν is a restriction of μ, $\Xi = \mathit{dom}(\Lambda_\mu)$ is a proper subset of $\mathit{dom}(\Lambda_\nu)$ and, more remarkably, no $a \in A$ is contained in $\mathit{aff}(\nu)$. Thus, in order to determine GMLE's for full exponential families in minimal representation in all cases of statistical interest, it is necessary to consider also non-full families with underlying measure concentrated on a proper affine subspace of $\mathbb{R}^d$ and a not contained in that subspace. We admit that no such intrinsic need is apparent for going beyond the case when $\mathit{dom}(\Lambda)$ has nonempty interior that contains some $\vartheta \in \Xi$, which will be called the *nondegenerate case.* We do not make that restriction mainly for mathematical completeness. This does lead to technical problems not present in the nondegenerate case, but these are not difficult to overcome via the concept of partial mean, see Subsection 2.4. Full generality is useful also for at least one application, viz. an explicit description of generalized rI-projections to canonically convex exponential families, even of pm's that do not have a mean, as indicated in Section 5.

Example 1.4. Let C be the closed convex cone in $\mathbb{R}^3$ given by $x_1 \geqslant 0$ and $x_1^2 \geqslant x_2^2 + x_3^2$. Let μ be sum of the unit point mass ν at the origin $c = (0, 0, 0)$ and of the pm concentrated on the boundary ∂C of C with the density $\frac{1}{2\pi} e^{-r} dr\, d\phi$, $r \geqslant 0$, $0 \leqslant \phi < 2\pi$, in cylindrical coordinates. The convex support of μ equals C, and for $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3)$ in the interior of $(1, 0, 0) - C$

$$\Lambda(\vartheta) = \ln\left[1 + \tfrac{1}{2\pi}\int_0^{2\pi} d\phi \int_0^{+\infty} e^{r(\vartheta_1 - 1 + \vartheta_2 \cos\phi + \vartheta_3 \sin\phi)}\, dr\right] = \ln\left[1 + \left[(\vartheta_1 - 1)^2 - \vartheta_2^2 - \vartheta_3^2\right]^{-\frac{1}{2}}\right]$$

and $\Lambda(\vartheta) = +\infty$ otherwise. Thus $\mathit{dom}(\Lambda)$ equals this shifted open cone, see Fig. 1.

Consider the full exponential family $\mathcal{E}_\mu$, thus let $\Xi = \mathit{dom}(\Lambda)$. If a is in the interior of C, elementary calculations show that the function $\vartheta \mapsto \langle\vartheta, a\rangle - \Lambda(\vartheta)$ has a unique maximizer $\vartheta^* \in \mathit{dom}(\Lambda)$, the pm Q_{ϑ^*} has the mean a and equals $R^*(a)$.

For $a = (a_1, a_2, a_3) \in \partial C$ the function has no maximizer but is bounded above. In fact, $\langle\vartheta, a\rangle - \Lambda(\vartheta) < \langle\vartheta, a\rangle \leqslant a_1$ for $\vartheta \in \mathit{dom}(\Lambda)$ whence $\Lambda^*(a) \leqslant a_1$ similarly to [6, Example 2]. Actually, $\Lambda^*(a) = a_1$ by limiting for $\vartheta_n = (1 - \frac{1}{n}, 0, 0) + n^2 b \in \mathit{dom}(\Lambda)$, $n \to \infty$, where $b \in -\partial C$ is nonzero and orthogonal to a. Since $\Lambda^*(a) = \Psi^*(a)$ is finite, the concept of

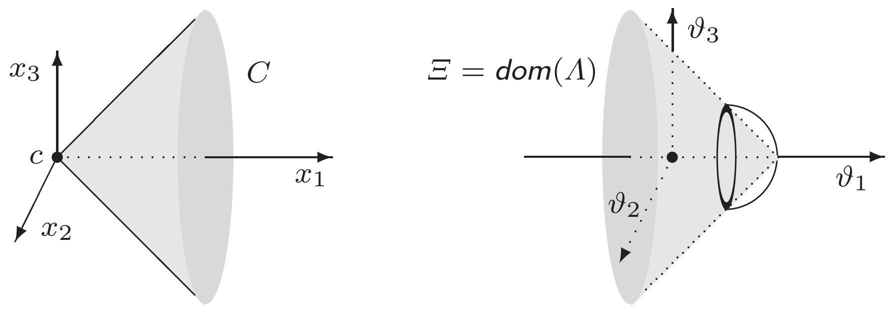

FIGURE 1.

GMLE applies, and (2) with $R^*(a)$ replaced by ν rewrites to

$$a_1 - \big[\langle\vartheta, a\rangle - \Lambda(\vartheta)\big] \geqslant D(\nu\|Q_\vartheta) = \Lambda(\vartheta)\,, \qquad \vartheta \in \Xi\,,$$

which obviously holds. Thus, $R^*_{\mu,\Xi}(a) = \nu$ by uniqueness of GMLE discussed in Remark 1.2.

On the other hand, $\Lambda_\nu = 0$ and the family $\mathcal{E}_{\nu,\Xi} = \{\nu\}$ is not in minimal representation. Obviously, $\Psi^*_{\nu,\Xi}(a)$ equals a_1 if $a \in C$ and $+\infty$ otherwise. In the cases $a \in \partial C$, the equalities $\Psi^*_{\mu,\Xi}(a) = \Psi^*_{\nu,\Xi}(a)$ and $R^*_{\nu,\Xi}(a) = R^*_{\mu,\Xi}(a)$ hold, and thus a characterization of $R^*(a)$ is indeed related to the maximization of $\langle\vartheta, a\rangle - \Lambda_\nu(\vartheta)$ subject to $\vartheta \in \Xi$, where $a \notin \mathit{aff}(\nu)$ unless $a = c$. These observations are not particular to this simple example but rather illustrate general phenomena studied in this paper.

For an i.i.d. sample of any size from any pm in the family $\mathcal{E}_\mu$, the probability is positive that exactly one element differs from c, thus the sample mean a belongs to $A = \partial C \setminus \{c\}$ with a positive probability. Example 5.5 of Section 5 exposes further illustrations of results of this paper for the full family $\mathcal{E}_\mu$.

1.3. The main concepts and notations are introduced in Section 2 including tools from convex geometry. Several auxiliary results are also collected there. Section 3 deals with the case when a belongs to a certain convex subset of $\mathit{dom}(\Psi^*)$. As will be clear later, this set contains the relative interior of $\mathit{dom}(\Psi^*)$ and is exactly the set of those $a \in \mathit{dom}(\Psi^*)$ for which $R^*(a)$ belongs to $\mathcal{E}$, see Corollary 4.2. For a in this subset, the supremum in (1) is actually a maximum, perhaps not subject to $\vartheta \in \Xi$ but to $\vartheta \in \Xi_{\mu,a}$ where the latter set is the closure of a projection of Ξ such that the values $\langle\vartheta, a\rangle - \Lambda(\vartheta)$, $\vartheta \in \mathit{dom}(\Lambda)$, are preserved by this projection. While this section is still of preparatory character, it has also independent interest because results well known under some regularity conditions [1, Section 9.4], are proved here in full generality. Much of the effort is needed to cover also the case $a \notin \mathit{aff}(\mu)$, disregarded in previous literature because of no immediate statistical meaning.

In Section 4, a pm satisfying (2), thus the GMLE from $a \in \mathit{dom}(\Psi^*)$, is constructed explicitly in Theorem 4.1, in contrast to [6] where $R^*(a)$ was obtained as the limit of a Cauchy sequence in a complete space. In the case not covered in Section 3, a proper face of the set called convex core of μ is identified by geometric means. For the family $\mathcal{E}_{\nu,\Xi}$ determined by the restriction ν of μ to the closure of this face, results of Section 3 apply to obtain $R^*_{\nu,\Xi}(a)$. This equals the desired GMLE $R^*_{\mu,\Xi}(a)$. Relying on this construction, also descriptions of $\mathit{dom}(\Psi^*)$ are presented in Theorem 4.9.

In Section 5, several properties of the GMLE mapping $R^*\colon a \mapsto R^*(a)$, defined on $\mathit{dom}(\Psi^*)$, are proved. Its range is characterized as a subset of $\mathit{cl}_v(\mathcal{E}_\Xi)$, Theorem 5.1. In the nondegenerate case, this subset consists exactly of those pm's that have a mean. The inverse of the GMLE mapping is also addressed in Theorem 5.1. For each pm P in the range, the set $\{a \in \mathit{dom}(\Psi^*) \colon R^*(a) = P\}$ is shown to be the singleton $\{\mathit{m}(P)\}$ consisting of the partial mean $\mathit{m}(P)$ of P or a (not necessarily convex) cone shifted by $\mathit{m}(P)$. A general result about the continuity of the GMLE mapping is presented, Theorem 5.6, and special situations are discussed where the GMLE mapping is bijective or even a homeomorphism, Remark 5.9. Then, a relationship of the concepts of GMLE and generalized rI-projection is discussed. Finally, the log-convexity of $\mathit{cl}_v(\mathcal{E}_\Xi)$ and of the range of R^* is established, Theorem 5.10, in the sense of [6].

In Section 6, an MLE in $\mathit{cl}_v(\mathcal{E}_\Xi)$ is considered. This MLE concept extends that in Subsection 1.1 when $a \in \mathit{aff}(\mu)$, but is undefined otherwise. It is shown that if this MLE exists then it corresponds to the GMLE, but not conversely. In the special case $\mathit{dom}(\Lambda) = \mathbb{R}^d$, the GMLE and MLE coincide, and the results of this paper on GMLE's cover previous ones on nonexistence of MLE's. Finally, in Appendix the invariance of the concepts of this paper w.r.t. different parameterizations of a family is treated. The GMLE $R^*(a)$ is shown not to depend on the parametrization when $a \in \mathit{aff}(\mu)$, though this invariance breaks down in other cases without statistical interpretation.

1.4. Exponential families are covered in depth in the monographs Chentsov [3], Barndorff-Nielsen [1], Brown [2] and Letac [11]. The standard reference for convexity is Rockafellar [15].

The ML estimation in canonically convex exponential families is treated in detail in [1, Section 9.4]. Generalizations of MLE's for the case when the supremum in (1) is not attained, have been considered for exponential families of pm's concentrated on finite or countable sets. The former case, in a theoretical sense, is completely settled in [1, Theorem 9.16], replacing the family by its "completion"; for recent related works addressing also computational issues in the framework of contingency tables see [10, 13]. The latter case is treated in [2, Chapter 6] via "aggregate families", under rather restrictive additional conditions. In general, one natural "completion" is the closure in variation distance. Some results on MLE in that direction, slightly improved here in Section 6, appeared in Csiszár and Matúš [6, Section 6]. The closure in variation distance of any canonically convex exponential family is characterized in Csiszár and Matúš [8], using the concepts of convex core of a measure introduced in [4], and of its accessible faces introduced in [8]. The log-convexity of this closure has not been yet addressed, but its subset consisting of all pm's to which some sequence in $\mathcal{E}_\Xi$ converges in rI-divergence (containing all GMLE's) has been shown to be not necessarily log-convex, Csiszár and Matúš [7].

A description of $\mathit{dom}(\Psi^*)$, even of $\mathit{dom}(\Lambda^*)$, has been elusive in general, though a relationship between them has been known under some conditions [1, p. 159, eq. (3)]. The strongest previous results on $\mathit{dom}(\Lambda^*)$ are apparently those in [6, Proposition 1].

The concept of GMLE, the main subject of this paper, is introduced in [5, 6], motivated by the study of generalized rI-projections. The nonconstructive existence proof of GMLE given in [6] is extended to infinite dimensional exponential families in [9]. A constructive proof of the existence of GMLE is first given in this paper.

One application area of GMLE's seems to be within graphical models of multivariate statistics, see [12, Chapter 6] where MLE's have been considered for various general exponential families which reduce by sufficiency to standard ones considered in this paper.

In such general mixed models, nonexistence of MLE's is a rather common phenomenon, and GMLE's may provide a remedy. In the particular case of log-linear models for multi-dimensional contingency tables computational strategies under the nonexistence of MLE are discussed in [14].

2. Preliminaries

2.1. Denote for a subset B of $\mathbb{R}^d$ its closure by $\mathit{cl}(B)$, affine hull by $\mathit{aff}(B)$, relative interior by $\mathit{ri}(B)$, which is the interior of B in the topology of $\mathit{aff}(B)$, and denote the shift of $\mathit{aff}(B)$ containing the origin by $\mathit{lin}(B)$, which is the linear space spanned by the differences $b-c$, $b,c\in B$. The orthogonal projector to $\mathit{lin}(B)$ is denoted by π_B. The orthogonal complement of a linear subspace E of $\mathbb{R}^d$ is denoted by $E^\perp$.

A *face* of a convex set $C\subseteq\mathbb{R}^d$ is a nonempty convex subset F that contains $ta+(1-t)b$ for some a,b in C and $0<t<1$ only if a and b are in F; note the slight deviation from the terminology of [15] where also the empty set is a face.

Lemma 2.1. *For convex subsets C and K of $\mathbb{R}^d$ and $a\in C+K$, in the family of faces F of C with $a\in \mathit{ri}(F)+K$ the inclusion-largest element exists.*

Proof. If a is in $\mathit{ri}(F_i)+K$ for faces F_i of C, $i=1,2$, thus $a=f_i+k_i$ with $f_i\in\mathit{ri}(F_i)$ and $k_i\in K$, then the element $g=\frac{1}{2}(f_1+f_2)$ of C belongs to $\mathit{ri}(G)$ for a unique face G of C [15, Theorem 18.2]. Since $a=g+\frac{1}{2}(k_1+k_2)$ it follows from convexity of K that $a\in\mathit{ri}(G)+K$. The definition of face implies that the segment with endpoints f_1 and f_2 is contained in G. Hence, both $\mathit{ri}(F_i)$ intersect G, and thus are contained in G [15, Theorem 18.1]. This implies that a face in the family with the largest dimension contains all faces of the family. □

A subset of $\mathbb{R}^d$ is a *cone* if it contains tx with each of its elements x and $t>0$. For a convex set $C\subseteq\mathbb{R}^d$, its *recession cone* $\mathit{rec}(C)$ is the set of all $y\in\mathbb{R}^d$ such that $x+ty\in C$ for all $x\in C$ and $t\geqslant 0$. In [15], the notation 0^+C is used instead of $\mathit{rec}(C)$.

Lemma 2.2. *If T is a linear mapping and $C\subseteq\mathbb{R}^d$ is convex then $T\mathit{rec}(C)\subseteq\mathit{rec}(TC)$.*

Proof. Given $a\in T\mathit{rec}(C)$ and $b\in TC$, write $a=Tc_a$, $b=Tc_b$ where $c_a\in\mathit{rec}(C)$, $c_b\in C$. Then $c_b+tc_a\in C$ for $t\geqslant 0$. Applying T, $b+ta\in TC$, and thus $a\in\mathit{rec}(TC)$. □

The *barrier cone* $\mathit{bar}(\Gamma)$ of a set $\Gamma\subseteq\mathbb{R}^d$ is the set of all $x\in\mathbb{R}^d$ such that the mapping $\vartheta\mapsto\langle\vartheta,x\rangle$ is bounded above on Γ. A vector $x\in\mathbb{R}^d$ is *normal* to the set Γ at $\theta\in\mathbb{R}^d$ if $\langle\vartheta-\theta,x\rangle\leqslant 0$ for all $\vartheta\in\Gamma$. Such a vector obviously belongs to $\mathit{bar}(\Gamma)$. The set of those normal vectors is the *normal cone* of Γ at θ, denoted by $N_\theta(\Gamma)$. In this paper, these terms are used in an extended sense, as the usual requirement $\theta\in\Gamma$ is not imposed.

Lemma 2.3. *For any nonempty convex $\Gamma\subseteq\mathbb{R}^d$,*

$$\mathit{rec}(\Gamma)\subseteq\mathit{rec}(\mathit{cl}(\Gamma))=\mathit{rec}(\mathit{ri}(\Gamma))=N_0(\mathit{bar}(\Gamma)).$$

Proof. The recession cone of $\mathit{cl}(\Gamma)$ contains that of Γ by [15, Theorem 8.3] and equals that of $\mathit{ri}(\Gamma)$ by [15, Corollary 8.3.1]. By [15, Corollary 14.2.1], where the normal cone at 0 of a cone K is called the polar of K, it holds that $\mathit{rec}(\mathit{cl}(\Gamma))=N_0(\mathit{bar}(\mathit{cl}(\Gamma)))$. This proves the last equality as $\mathit{bar}(\mathit{cl}(\Gamma))=\mathit{bar}(\Gamma)$ is obvious by definition. □

2.2. A face F of a convex set $C \subseteq \mathbb{R}^d$ is *exposed* if either $F = C$ or, for some unit vector τ, the maximum of $\langle \tau, x \rangle$ subject to $x \in C$ is attained if and only if $x \in F$. Such τ is said to *expose* F in C.

The following concepts were introduced in [8] in order to characterize the variation closure of $\mathcal{E}_{\Xi}$. An *access sequence* to a proper face F of a convex set $C \subseteq \mathbb{R}^d$ is an orthonormal sequence $\tau_1, \ldots, \tau_m$ such that $\tau_i \in \mathit{lin}(F_{i-1})$ exposes a face F_i of F_{i-1} for $i = 1, \ldots, m$, where $F_0 = C$ and $F_m = F$. Such a sequence always exists. In particular, an access sequence of length $m = 1$ to a proper face F exists if and only if F is an exposed face. An access sequence to F is *adapted* to a convex set $\Xi \subseteq \mathbb{R}^d$ if $\tau_i \in \mathit{rec}(\pi_{F_{i-1}}(\mathit{ri}(\Xi)))$ for $1 \leqslant i \leqslant m$. The access sequence to $F = C$ is empty, by definition, and it is adapted to any convex Ξ. Note that the notion of adaptedness depends on Ξ only through $\mathit{ri}(\Xi)$ or $\pi_C(\Xi)$ or $\pi_C(\mathit{ri}(\Xi)) = \mathit{ri}(\pi_C(\Xi))$. A face F of C is *Ξ-accessible* if there exists an access sequence to F which is adapted to Ξ. More details and examples illustrating these notions can be found in [8].

Lemma 2.4. *For intersecting faces F and G of a convex set C, if F is a Ξ-accessible face of C then $F \cap G$ is a Ξ-accessible face of G.*

Proof. The case $F \supseteq G$ is trivial. Otherwise, $F \neq C$ and there exists an access sequence $\tau_1, \ldots \tau_m$ to the face F of C which is adapted to Ξ. Let $C = F_0 \supset \ldots \supset F_m = F$ be the corresponding chain of faces, in particular, $\tau_i \in \mathit{rec}(\pi_{F_{i-1}}(\mathit{ri}(\Xi)))$, $1 \leqslant i \leqslant m$.

The sequence $G_i = F_i \cap G$ of faces of G decreases from G to $F \cap G$ and τ_i exposes G_i in G_{i-1}. If G_{i-1} contains G_i strictly then τ_i is not orthogonal to $\mathit{lin}(G_{i-1})$ and the unit vector ϑ_i in the direction $\pi_{G_{i-1}}(\tau_i)$ exposes G_i in G_{i-1} as well. This ϑ_i belongs to $\pi_{G_{i-1}}(\mathit{rec}(\pi_{F_{i-1}}(\mathit{ri}(\Xi))))$, thus to $\mathit{rec}(\pi_{G_{i-1}}(\mathit{ri}(\Xi)))$ by Lemma 2.2 and $\pi_{G_{i-1}}\pi_{F_{i-1}} = \pi_{G_{i-1}}$.

Let $i_1 < \ldots < i_k$ be those indices $1 \leqslant i \leqslant m$ for which G_i is a proper face of G_{i-1}. For $1 \leqslant j \leqslant k$ the face G_{i_j-1} of G equals $G_{i_{j-1}}$, and hence ϑ_{i_j} exposes G_{i_j} in $G_{i_{j-1}}$ and belongs to $\mathit{rec}(\pi_{G'}(\mathit{ri}(\Xi)))$ where $G' = G_{i_{j-1}}$. Thus, $\vartheta_1, \ldots \vartheta_k$ is an access sequence to $F \cap G$ from G, adapted to Ξ. $\square$

Corollary 2.5. *If F and G are intersecting Ξ-accessible faces of a convex set C then $F \cap G$ is also a Ξ-accessible face of C.*

Proof. Since G is Ξ-accessible from C and $F \cap G$ is Ξ-accessible from G by Lemma 2.4 it suffices to concatenate their access sequences adapted to Ξ. The concatenation is an access sequence to the face $F \cap G$ of C, adapted to Ξ. $\square$

2.3. For a σ-finite Borel measure μ on $\mathbb{R}^d$, its *convex support* $\mathit{cs}(\mu)$ and *convex core* $\mathit{cc}(\mu)$ are defined as intersections of those convex sets $C \subseteq \mathbb{R}^d$ that are μ-full, $\mu(\mathbb{R}^d \setminus C) = 0$, and closed, respectively Borel. The cores have been considered previously only for finite measures. Since $\mathit{cs}(\mu)$ and $\mathit{cc}(\mu)$ do not change when μ is replaced by any equivalent measure, results on finite measures immediately extend to σ-finite ones (provided that when images of σ-finite measures are considered, these are postulated to be σ-finite). References are made below to the following assertions.

Fact 2.6. *The closure of $\mathit{cc}(\mu)$ equals $\mathit{cs}(\mu)$* [4, Lemma 1]*, thus $\mathit{cc}(\mu)$ is nonempty if and only if μ is nonzero.*

Fact 2.7. *Every face F of $\mathsf{cc}(\mu)$ equals the convex core of the restriction of μ to $\mathsf{cl}(F)$* [4, Lemma 3], *hence this restriction is nonzero.*

Fact 2.8. *The image of $\mathsf{cc}(\mu)$ under any affine transformation T equals the convex core of the image of μ under T* [4, Lemma 8], *providing that image is σ-finite.*

Fact 2.9. *A supporting hyperplane H of $\mathsf{cs}(\mu)$ has positive μ-measure if and only if $F = H \cap \mathsf{cc}(\mu)$ is nonempty, in which case $\mu(H \setminus \mathsf{cl}(F)) = 0$* [8, Lemma 1].

Fact 2.10. *If F and G are faces of $\mathsf{cc}(\mu)$ then $\mathsf{cl}(F) \cap \mathsf{cl}(G) \setminus \mathsf{cl}(F \cap G)$ has μ-measure zero* [4, Corollary 4].

On account of Fact 2.6, $\mathsf{cc}(\mu)$ and $\mathsf{cs}(\mu)$ have the same relative interior, denoted by $\mathsf{ri}(\mu)$. The notations $\mathsf{aff}(\mu)$, already used in Section 1, and $\mathsf{lin}(\mu)$ are analogous and selfexplaining.

For $a \in \mathbb{R}^d$ let $E_{\mu,a}$ denote the subspace of $\mathbb{R}^d$ spanned by the set $\{b - a\colon b \in \mathsf{aff}(\mu)\}$, or equivalently by $\mathsf{lin}(\mu)$ and $b - a$ for any fixed $b \in \mathsf{aff}(\mu)$. Thus, $E_{\mu,a} \supseteq \mathsf{lin}(\mu)$, with equality if and only if $a \in \mathsf{aff}(\mu)$. Let π_μ and $\pi_{\mu,a}$ denote the orthogonal projectors to $\mathsf{lin}(\mu)$ and $E_{\mu,a}$, respectively.

Lemma 2.11. *Each $\vartheta \in \mathbb{R}^d$ satisfies $\langle \vartheta, a\rangle - \Lambda(\vartheta) = \langle \pi_{\mu,a}(\vartheta), a\rangle - \Lambda(\pi_{\mu,a}(\vartheta))$.*

Proof. If $x \in \mathsf{aff}(\mu)$ then $x - a \in E_{\mu,a}$, thus $\langle \vartheta, x - a\rangle = \langle \pi_{\mu,a}(\vartheta), x - a\rangle$. Hence,

$$\langle \vartheta, a\rangle - \Lambda(\vartheta) = -\ln \int_{\mathsf{aff}(\mu)} e^{\langle \vartheta, x-a\rangle}\, \mu(dx)$$

does not change when ϑ is replaced by $\pi_{\mu,a}(\vartheta)$. $\square$

Recall the standing convention that $\varXi$ denotes a nonempty convex subset of $\mathsf{dom}(\Lambda)$ and $\Psi^* = \Psi^*_{\mu,\varXi}$ is defined by (1).

Corollary 2.12. $\Psi^*_{\mu,\varXi}(a) = \Psi^*_{\mu,\pi_{\mu,a}(\varXi)}(a)$. *If $\Gamma \subseteq \mathsf{dom}(\Lambda)$ and $\pi_{\mu,a}(\Gamma) = \pi_{\mu,a}(\varXi)$ then $\Psi^*_{\mu,\Gamma}(a) = \Psi^*_{\mu,\varXi}(a)$.*

The special instance of Lemma 2.11

$$\langle \vartheta, b\rangle - \Lambda(\vartheta) = \langle \pi_\mu(\vartheta), b\rangle - \Lambda(\pi_\mu(\vartheta))\,, \qquad \vartheta \in \mathbb{R}^d\,,\ b \in \mathsf{aff}(\mu)\,, \tag{3}$$

is well-known. Its consequences include the facts that Q_ϑ with $\vartheta \in \mathsf{dom}(\Lambda)$ does not change when ϑ is replaced by $\pi_\mu(\vartheta)$, and $\mathsf{dom}(\Lambda) = \pi_\mu(\mathsf{dom}(\Lambda)) + \mathsf{lin}(\mu)^\perp$.

Lemma 2.13. $\Psi^*_{\mu,\mathsf{ri}(\varXi)}(a) = \Psi^*_{\mu,\varXi}(a) = \Psi^*_{\mu,\mathsf{cl}(\varXi)\cap\mathsf{dom}(\Lambda)}(a)$.

Proof. If $\theta \in \mathsf{cl}(\varXi) \cap \mathsf{dom}(\Lambda)$ then some sequence ϑ_n in $\mathsf{ri}(\varXi)$ converges to θ along a segment. By continuity of Λ along the segment, $\Psi^*_{\mu,\mathsf{ri}(\varXi)}(a) \geqslant \langle \theta, a\rangle - \Lambda(\theta)$. Hence, $\Psi^*_{\mu,\mathsf{ri}(\varXi)}(a) \geqslant \Psi^*_{\mu,\mathsf{cl}(\varXi)\cap\mathsf{dom}(\Lambda)}(a)$. The remaining inequalities are trivial. $\square$

2.4. For any pm P on $\mathbb{R}^d$, write

$$\mathsf{M}(P) = \{\tau \in \mathbb{R}^d\colon x \mapsto \langle \tau, x\rangle \text{ is } P\text{-integrable}\}$$

and define the *partial mean* $\mathsf{m}(P)$ of P as the unique element of the linear space $\mathsf{M}(P)$ that satisfies

$$\int_{\mathbb{R}^d} \langle \tau, x\rangle\, P(dx) = \langle \tau, \mathsf{m}(P)\rangle\,, \quad \tau \in \mathsf{M}(P)\,.$$

Note that $\mathsf{M}(P) = \mathbb{R}^d$ if and only if P has a mean, in which case $\mathsf{m}(P)$ equals the mean of P.

Lemma 2.14. *For any pm P on $\mathbb{R}^d$, $\mathsf{lin}(P)^{\perp}$ is a subspace of $\mathsf{M}(P)$.*

Proof. If $\tau \in \mathsf{lin}(P)^{\perp}$ then $x \mapsto \langle \tau, x\rangle$ is constant on $\mathsf{aff}(P)$, and thus $\tau \in \mathsf{M}(P)$. □

Lemma 2.15. *The partial mean $\mathsf{m}(P)$ of a pm P belongs to $\mathsf{ri}(P) + \mathsf{M}(P)^{\perp}$, contained in $\mathsf{aff}(P)$.*

Proof. Let π denote, in this proof only, the orthogonal projector to $\mathsf{M}(P)$, and πP the image of P under π. Then for $\vartheta \in \mathsf{M}(P)$

$$\int_{\mathbb{R}^d} \langle \vartheta, x\rangle \,(\pi P)(dx) = \int_{\mathbb{R}^d} \langle \vartheta, \pi(x)\rangle \, P(dx) = \int_{\mathbb{R}^d} \langle \vartheta, x\rangle \, P(dx) = \langle \vartheta, \mathsf{m}(P)\rangle$$

while for $\vartheta \in \mathsf{M}(P)^{\perp}$ clearly

$$\int_{\mathbb{R}^d} \langle \vartheta, x\rangle \,(\pi P)(dx) = 0 = \langle \vartheta, \mathsf{m}(P)\rangle \,.$$

Hence, $\mathsf{M}(\pi P) = \mathbb{R}^d$ and the mean of πP exists and equals $\mathsf{m}(P)$. It follows that $\mathsf{m}(P)$ belongs to the relative interior of $\mathsf{cs}(\pi P)$, and then by Fact 2.6 to

$$\mathsf{ri}(\mathsf{cc}(\pi P)) = \mathsf{ri}(\pi\,\mathsf{cc}(P)) = \pi(\mathsf{ri}(\mathsf{cc}(P)))\,.$$

Here, the interchangeability of cc and the projector holds by Fact 2.8 and that of relative interior and projectors is obvious. This implies that $\mathsf{m}(P)$ is in $\pi(\mathsf{ri}(P))$, and thus in $\mathsf{ri}(P) + \mathsf{M}(P)^{\perp}$. This is a subset of $\mathsf{aff}(P)$ because $\mathsf{ri}(P) \subseteq \mathsf{aff}(P)$ and $\mathsf{M}(P)^{\perp} \subseteq \mathsf{lin}(P)$, the latter by Lemma 2.14. □

The relevance of partial means for exponential families is indicated by the following simple lemma, particularly by its special instance $P = Q_\theta$.

Lemma 2.16. *For ϑ and θ in $\mathsf{dom}(\Lambda)$ and a pm P with $D(P\|Q_\theta) < \infty$, the necessary and sufficient condition for $D(P\|Q_\vartheta) < \infty$ is $\vartheta - \theta \in \mathsf{M}(P)$. Under that condition*

$$D(P\|Q_\vartheta) - D(P\|Q_\theta) = \langle \theta - \vartheta, \mathsf{m}(P)\rangle - \Lambda(\theta) + \Lambda(\vartheta)\,. \tag{4}$$

Proof. The left-hand side of (4) is, by definition, the difference of two integrals. If one of them is finite, the difference can be written as one integral which is equal to $\int \langle \theta - \vartheta, x\rangle \, P(dx) - \Lambda(\theta) + \Lambda(\vartheta)$. This is finite if and only if $\vartheta - \theta \in \mathsf{M}(P)$, in which case it equals the right-hand side of (4). □

Corollary 2.17. *The I-divergence $D(P\|Q_\vartheta)$ is finite for each $\vartheta \in \Xi$ if and only if it is finite for one $\vartheta \in \Xi$ and $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(P)$.*

The instance $P = Q_\theta$ of Lemma 2.16 is that the I-divergence $D(Q_\theta\|Q_\vartheta)$ is finite if and only if $\vartheta - \theta$ belongs to $\mathsf{M}(Q_\theta)$, in which case

$$D(Q_\theta\|Q_\vartheta) = \langle \theta - \vartheta, \mathsf{m}(Q_\theta)\rangle - \Lambda(\theta) + \Lambda(\vartheta)\,. \tag{5}$$

The corresponding instance of Corollary 2.17 is that if $\theta \in \Xi$ then $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$ is equivalent to the finiteness of $D(Q_\theta\|Q_\vartheta)$ for each $\vartheta \in \Xi$. To extend this equivalence to θ beyond Ξ, let

$$\Xi_\mu = \mathsf{cl}(\pi_\mu(\Xi)) \cap \mathsf{dom}(\Lambda_\mu) \quad \text{and} \quad \tilde{\Xi}_\mu = \Xi_\mu + \mathsf{lin}(\mu)^{\perp}\,,$$

and observe that $\Xi \subseteq \tilde{\Xi}_\mu$.

Lemma 2.18. *For $\theta \in \tilde{\Xi}_\mu$, the inclusion $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$ is equivalent to $\vartheta - \theta \in \mathsf{M}(Q_\theta)$ for each $\vartheta \in \Xi$, or to the finiteness of $D(Q_\theta\|Q_\vartheta)$ for each $\vartheta \in \Xi$.*

Proof. Since clearly $\mathsf{lin}(\Xi_\mu) = \mathsf{lin}(\pi_\mu(\Xi))$, and $\mathsf{M}(Q_\theta)$ contains $\mathsf{lin}(Q_\theta)^\perp = \mathsf{lin}(\mu)^\perp$ by Lemma 2.14, the condition $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$ is equivalent to $\mathsf{lin}(\Xi_\mu) \subseteq \mathsf{M}(Q_\theta)$. On account of $\pi_\mu(\theta) \in \Xi_\mu$ the latter is, in turn, equivalent to $\pi_\mu(\vartheta) - \pi_\mu(\theta) \in \mathsf{M}(Q_\theta)$, $\vartheta \in \Xi$, and using Lemma 2.14 again, to $\vartheta - \theta \in \mathsf{M}(Q_\theta)$, $\vartheta \in \Xi$. The second equivalence follows from Lemma 2.16 with $P = Q_\theta$. □

Lemma 2.19. *If Ξ intersects $\mathsf{ri}(\mathsf{dom}(\Lambda))$ and $\theta \in \tilde{\Xi}_\mu$ then $\mathsf{M}(Q_\theta)$ contains $\mathsf{lin}(\Xi)$ if and only if it contains $\mathsf{lin}(\mathsf{dom}(\Lambda))$.*

Proof. By Lemma 2.18, the second assumption and $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$ imply that Ξ is contained in $[\theta + \mathsf{M}(Q_\theta)] \cap \mathsf{dom}(\Lambda)$. This set is a face of $\mathsf{dom}(\Lambda)$ due to Lemma 2.21 below. The first assumption implies that this face cannot be proper. Hence, $\theta + \mathsf{M}(Q_\theta)$ contains $\mathsf{dom}(\Lambda)$, and thus the assertion follows. □

Corollary 2.20. *In the nondegenerate case, when Ξ intersects the interior of $\mathsf{dom}(\Lambda)$, a pm Q_θ with $\theta \in \tilde{\Xi}_\mu$ has a mean if and only if $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$.*

The instance $\Xi = \{\theta\}$ of this corollary covers the well-known fact that the pm Q_θ indexed by an interior point θ of $\mathsf{dom}(\Lambda)$ has a mean.

Lemma 2.21. *If $\theta \in \mathsf{dom}(\Lambda)$ then $[\theta + \mathsf{M}(Q_\theta)] \cap \mathsf{dom}(\Lambda)$ is a face of $\mathsf{dom}(\Lambda)$.*

Proof. For $\vartheta, \tau \in \mathsf{dom}(\Lambda)$ and $0 < t < 1$

$$t\int_{\mathbb{R}^d} \langle \vartheta - \theta, x\rangle\, Q_\theta(dx) + (1-t)\int_{\mathbb{R}^d} \langle \tau - \theta, x\rangle\, Q_\theta(dx) = \int_{\mathbb{R}^d} \langle t\vartheta + (1-t)\tau - \theta, x\rangle\, Q_\theta(dx)$$

where each integral is finite or $-\infty$ by [8, Lemma 4]. If $\Gamma = [\theta + \mathsf{M}(Q_\theta)] \cap \mathsf{dom}(\Lambda)$ contains $t\vartheta + (1-t)\tau$ then the integral on the right is finite, and hence all integrals are finite. Thus, $\vartheta, \tau \in \Gamma$, and Γ is a face. □

2.5. The *subdifferential* at $\theta \in \mathbb{R}^d$ of a convex function f on $\mathbb{R}^d$ is the set $\partial f(\theta)$ consisting of those $x \in \mathbb{R}^d$ that satisfy $f(\vartheta) \geqslant f(\theta) + \langle \vartheta - \theta, x\rangle$ for all $\vartheta \in \mathbb{R}^d$ or, equivalently, $f^*(x) = \langle \theta, x\rangle - f(\theta)$, where f^* is the convex conjugate of f. Subdifferentials appear in this paper only when comparing results with previous ones. In the following lemma, Ψ denotes the function defined in Remark 1.1 that equals Λ on the convex nonempty subset Ξ of $\mathsf{dom}(\Lambda)$ and $+\infty$ elsewhere.

Lemma 2.22. *If Ξ intersects $\mathsf{ri}(\mathsf{dom}(\Lambda))$ then $\mathsf{dom}(\Psi^*) = \mathsf{dom}(\Lambda^*) + \mathsf{bar}(\Xi)$ and*

$$\partial\Psi(\theta) = \begin{cases} \partial\Lambda(\theta) + \mathsf{N}_\theta(\Xi)\,, & \theta \in \Xi\,, \\ \emptyset\,, & \textit{otherwise.} \end{cases}$$

Remark 2.23. Analogous results appear in [1, Section 9.4], see (3) and (4) there, under the assumption that Ξ and $\mathsf{dom}(\Lambda)$ have nonempty interiors. Later, the first equality will be shown to hold for any convex nonempty $\Xi \subset \mathsf{dom}(\Lambda)$, see Remark 4.10 in Section 4.

Proof. The function Ψ is the sum of Λ and the indicator function $\delta(\cdot|\Xi)$ of Ξ, equal to 0 on Ξ and $+\infty$ elsewhere. As Ξ is contained in $\mathsf{dom}(\Lambda)$ but not in its relative boundary $\mathsf{dom}(\Lambda) \setminus \mathsf{ri}(\mathsf{dom}(\Lambda))$, the relative interior of Ξ is contained in that of $\mathsf{dom}(\Lambda)$ [15, Corollary 6.5.2]. Then, the first equality follows from [15, Theorem 16.4] using that $\mathsf{dom}(\delta^*(\cdot|\Xi))$ equals $\mathsf{bar}(\Xi)$, and the second one from [15, Theorem 23.8] since the subdifferential of $\delta(\cdot|\Xi)$ at θ equals $\mathsf{N}_\theta(\Xi)$ if $\theta \in \Xi$ and $\emptyset$ otherwise. □

3. The gmle when $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$

3.1. Given a nonempty convex set Ξ contained in $\mathit{dom}(\Lambda)$ and $a \in \mathbb{R}^d$, let

$$\Xi_{\mu,a} = \mathit{cl}(\pi_{\mu,a}(\Xi)) \cap \mathit{dom}(\Lambda) \quad \text{and} \quad \tilde{\Xi}_{\mu,a} = \Xi_{\mu,a} + E_{\mu,a}^{\perp},$$

see Subsection 2.3 for notation. These are convex subsets of $\mathit{dom}(\Lambda)$, and $\Xi \subseteq \tilde{\Xi}_{\mu,a}$. Then

$$\Psi^*_{\mu,\Xi}(a) = \Psi^*_{\mu,\Xi_{\mu,a}}(a) = \Psi^*_{\mu,\tilde{\Xi}_{\mu,a}}(a) \tag{6}$$

where the first equality follows from Corollary 2.12 and Lemma 2.13, and the second one from Corollary 2.12 applied to $\tilde{\Xi}_{\mu,a}$ in the role of Ξ. If $a \in \mathit{aff}(\mu)$ then $\Xi_{\mu,a}$ is equal to $\Xi_\mu = \mathit{cl}(\pi_\mu(\Xi)) \cap \mathit{dom}(\Lambda)$ and $\tilde{\Xi}_{\mu,a}$ to $\tilde{\Xi}_\mu$, see Subsection 2.4, while always $\pi_\mu(\Xi_{\mu,a}) \subseteq \Xi_\mu$, and hence $\tilde{\Xi}_{\mu,a} \subseteq \tilde{\Xi}_\mu$. In the sequel, notations introduced in Section 2 are used without reference.

The following theorem admits to define a mapping that assigns to each element a of $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$ the parameter $\theta^*(a) = \theta^*_{\mu,\Xi}(a)$ equal to the unique maximizer of the function $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$ subject to $\vartheta \in \Xi_{\mu,a}$. Necessity of the assumption $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$ for the existence of such a maximizer is shown in Theorem 3.2.

Theorem 3.1. *If $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$ then $\Psi^*(a)$ is finite and a unique $\theta \in \Xi_{\mu,a}$ exists such that $\Psi^*(a) = \langle \theta, a \rangle - \Lambda(\theta)$.*

Proof. If a belongs to $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$ then $a - b \in \mathit{bar}(\Xi)$ for some $b \in \mathit{ri}(\mu)$, thus

$$\langle \vartheta, a - b \rangle \leqslant r, \qquad \vartheta \in \Xi, \tag{7}$$

for some $r \in \mathbb{R}$. By (3) and [8, Lemma 9] applied to this b, for some $s > 0$ and $t \in \mathbb{R}$

$$\langle \vartheta, b \rangle - \Lambda(\vartheta) = \langle \pi_\mu(\vartheta), b \rangle - \Lambda(\pi_\mu(\vartheta)) \leqslant t - s\,\|\pi_\mu(\vartheta)\|, \qquad \vartheta \in \mathbb{R}^d. \tag{8}$$

Note that though Lemma 9 of [8] is formulated only for finite measures, its proof holds verbatim also for σ-finite measures μ with $\mathit{dom}(\Lambda_\mu)$ nonempty. Combining (7) and (8),

$$\langle \vartheta, a \rangle - \Lambda(\vartheta) - r \leqslant \langle \vartheta, b \rangle - \Lambda(\vartheta) \leqslant t - s\,\|\pi_\mu(\vartheta)\|, \qquad \vartheta \in \Xi, \tag{9}$$

whence $\Psi^*(a) \leqslant t + r$.

Consider a sequence ϑ_n in Ξ with $\langle \vartheta_n, a \rangle - \Lambda(\vartheta_n)$ converging to the finite number $\Psi^*(a)$. By (8) and (7)

$$\langle \vartheta_n, a \rangle - \Lambda(\vartheta_n) - t \leqslant \langle \vartheta_n, a - b \rangle \leqslant r.$$

Thus the sequence $\langle \vartheta_n, a - b \rangle$ is bounded, and by (9) the sequence $\pi_\mu(\vartheta_n)$ is bounded, too. As $\mathit{lin}(\mu)$ and $b - a$ span $E_{\mu,a}$, it follows that the sequence $\pi_{\mu,a}(\vartheta_n)$ is bounded. Hence, going to a subsequence if necessary, $\pi_{\mu,a}(\vartheta_n)$ converges to some $\theta \in \mathit{cl}(\pi_{\mu,a}(\Xi))$. By Lemma 2.11, $\langle \pi_{\mu,a}(\vartheta_n), a \rangle - \Lambda(\pi_{\mu,a}(\vartheta_n))$ converges to $\Psi^*(a)$. Then, $\Psi^*(a) \leqslant \langle \theta, a \rangle - \Lambda(\theta)$ by semicontinuity of Λ. In particular, $\theta \in \mathit{dom}(\Lambda)$, thus $\theta \in \Xi_{\mu,a}$. The opposite inequality follows from (6).

To prove uniqueness, it suffices to show that $\theta, \tau \in \Xi_{\mu,a}$ and

$$\Psi^*(a) = \langle \theta, a \rangle - \Lambda(\theta) = \langle \tau, a \rangle - \Lambda(\tau)$$

imply $\theta = \tau$. By (6), both θ and τ maximize the function $\vartheta \mapsto \langle \vartheta, a \rangle - \Lambda(\vartheta)$ over $\Xi_{\mu,a}$, and hence Λ is not strictly convex on the segment connecting θ and τ. It follows from (3) and the strict convexity of Λ on $\mathit{lin}(\mu)$ that $\pi_\mu(\theta)$ and $\pi_\mu(\tau)$ coincide. This and (3) imply that if $b \in \mathit{aff}(\mu)$ then $\langle \theta, b \rangle - \Lambda(\theta)$ equals $\langle \tau, b \rangle - \Lambda(\tau)$, and hence $\langle \theta, a - b \rangle = \langle \tau, a - b \rangle$. Since $\theta, \tau \in \Xi_{\mu,a} \subseteq E_{\mu,a}$, it follows that $\theta = \tau$. ☐

Theorem 3.2. *For any $a \in \mathbb{R}^d$ and $\theta \in \tilde{\Xi}_{\mu,a}$ the following statements are equivalent.*

(i) $\Psi^*(a) = \langle \theta, a\rangle - \Lambda(\theta)$.

(ii) $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$ *and* $a - \mathsf{m}(Q_\theta)$ *belongs to* $\mathsf{N}_\theta(\Xi)$.

(iii) $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ *and* $\theta^*(a) = \pi_{\mu,a}(\theta)$.

(iv) $\langle \theta, a\rangle - \Lambda(\theta) \geqslant \langle \vartheta, a\rangle - \Lambda(\vartheta) + D(Q_\theta \| Q_\vartheta)$ *for all* $\vartheta \in \Xi$.

By Theorem 3.2, in case $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ the GMLE $R^*(a)$ exists and equals $Q_{\theta^*(a)}$. Indeed, $\theta = \theta^*(a)$ satisfies *(iii)*, then *(i)* and *(iv)* combine to the GMLE inequality (2) with $R^*(a)$ replaced by $Q_{\theta^*(a)}$, and then $R^*(a) = Q_{\theta^*(a)}$ by uniqueness of GMLE, discussed in Remark 1.2. This result is complemented in Section 4, where Corollary 4.2 shows that if $a \notin \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ and the GMLE $R^*(a)$ exists then $R^*(a) \notin \mathcal{E}$.

Note that in the nondegenerate case, the inclusion in *(ii)* holds if and only if Q_θ has a mean, due to Corollary 2.20 and $\tilde{\Xi}_{\mu,a} \subseteq \tilde{\Xi}_\mu$. If, in addition, θ is in the interior of Ξ then the normal cone $\mathsf{N}_\theta(\Xi)$ is the singleton $\{0\}$, and the condition *(ii)* requires the mean of Q_θ to equal a.

Proof. (i)⇒(ii): On account of [8, Lemma 4], if $\theta \in \mathsf{dom}(\Lambda)$ and τ is a unit vector such that $\theta + t\tau \in \mathsf{dom}(\Lambda)$ for some $t > 0$ then $\int \langle \tau, x\rangle \, Q_\theta(dx)$ exists, either finite or $-\infty$, and is equal to the one-sided directional derivative of Λ at θ in the direction τ. The assumption *(i)* and (6) imply $\langle \theta, a\rangle - \Lambda(\theta) \geqslant \langle \vartheta, a\rangle - \Lambda(\vartheta)$ for all $\vartheta \in \tilde{\Xi}_{\mu,a}$. Hence, the derivative of the function $\vartheta \mapsto \langle \vartheta, a\rangle - \Lambda(\vartheta)$ at θ in any direction pointing from θ to some $\vartheta \in \tilde{\Xi}_{\mu,a}$ is not positive. It follows that for all $\vartheta \in \tilde{\Xi}_{\mu,a}$, the integral $\int \langle \vartheta - \theta, x\rangle \, Q_\theta(dx)$ is finite, thus $\vartheta - \theta \in \mathsf{M}(Q_\theta)$, and

$$0 \geqslant \langle \vartheta - \theta, a\rangle - \int_{\mathbb{R}^d} \langle \vartheta - \theta, x\rangle \, Q_\theta(dx) = \langle \vartheta - \theta, a - \mathsf{m}(Q_\theta)\rangle \,.$$

This means that $\mathsf{lin}(\tilde{\Xi}_{\mu,a})$ is contained in $\mathsf{M}(Q_\theta)$ and $a - \mathsf{m}(Q_\theta)$ is normal to $\tilde{\Xi}_{\mu,a}$ at θ. Since $\Xi \subseteq \tilde{\Xi}_{\mu,a}$ the validity of *(ii)* follows.

(i),(ii)⇒(iii): If $a - m(Q_\theta)$ belongs to $\mathsf{N}_\theta(\Xi) \subseteq \mathsf{bar}(\Xi)$ then $a \in m(Q_\theta) + \mathsf{bar}(\Xi)$. The assumption $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)$ implies, using Lemma 2.14, that $\mathsf{M}(Q_\theta)^\perp \subseteq \mathsf{lin}(\Xi)^\perp \subseteq \mathsf{bar}(\Xi)$. By this and $\mathsf{ri}(Q_\theta) = \mathsf{ri}(\mu)$, Lemma 2.15 gives that $\mathsf{m}(Q_\theta) \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$. As $\mathsf{bar}(\Xi)$ is a convex cone, $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ follows. Then $\theta^*(a) = \pi_{\mu,a}(\theta)$ is a consequence of *(i)* and Lemma 2.11, due to the uniqueness assertion of Theorem 3.1.

(iii)⇒(i): By the definition of $\theta^*(a)$, *(iii)* implies that the equality in *(i)* holds for $\pi_{\mu,a}(\theta)$ in the role of θ. Then, by Lemma 2.11, it holds also for θ.

(ii)⇔(iv): Due to $\theta \in \tilde{\Xi}_{\mu,a} \subseteq \tilde{\Xi}_\mu$ and Lemma 2.18, the inclusion in *(ii)* is equivalent to the finiteness of $D(Q_\theta \| Q_\vartheta)$ for each $\vartheta \in \Xi$. Then the second condition in *(ii)*, thus $0 \geqslant \langle \vartheta - \theta, a - \mathsf{m}(Q_\theta)\rangle$, $\vartheta \in \Xi$, is equivalent to *(iv)* on account of (5).

(iv)⇒(i): This obtains by taking supremum over $\vartheta \in \Xi$ in *(iv)*, where the I-divergence is nonnegative, and by (6). ☐

3.2. Theorem 3.2 makes it possible to describe the range of the mapping θ^*, based on the observation that the necessary and sufficient conditions for $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ and $\theta^*(a) = \theta$ are

$$\theta \in \Xi_{\mu,a}\,, \quad \mathsf{lin}(\mu) \subseteq \mathsf{M}(Q_\theta) \ \text{ and } \ a - \mathsf{m}(Q_\theta) \in \mathsf{N}_\theta(\Xi)\,. \tag{10}$$

Let

$$\Xi_\mu^{\mathsf{M}} = \{\theta \in \Xi_\mu \colon \mathit{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta)\}\,,$$

let $\boldsymbol{E}$ denote the family of linear subspaces E of $\mathbb{R}^d$ that contain $\mathit{lin}(\mu)$ as a subspace of codimension 1, and for $E \in \boldsymbol{E}$ denote

$$\Xi_E^{\mathsf{M},\cap} = \{\theta \in \mathit{cl}(\pi_E(\Xi)) \cap \mathit{dom}(\Lambda) \colon \mathit{lin}(\Xi) \subseteq \mathsf{M}(Q_\theta) \text{ and } \mathsf{N}_\theta(\Xi) \cap [E \setminus \mathit{lin}(\mu)] \neq \emptyset\}.$$

Lemma 3.3. *The range $\theta^*(\mathit{ri}(\mu) + \mathit{bar}(\Xi))$ of θ^* is equal to $\Xi_\mu^{\mathsf{M}} \cup \bigcup_{E\in\boldsymbol{E}} [\Xi_E^{\mathsf{M},\cap} \setminus \mathit{lin}(\mu)]$. For θ in this range, $\{a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi) \colon \theta^*(a) = \theta\}$ equals $\mathsf{m}(Q_\theta) + \mathsf{K}_\mu(\theta)$ where*

$$\mathsf{K}_\mu(\theta) = \mathsf{N}_\theta(\Xi) \cap \begin{cases} \mathit{lin}(\mu) \cup \bigcup \{E \in \boldsymbol{E} \colon \theta \in \Xi_E^{\mathsf{M},\cap}\}\,, & \theta \in \Xi_\mu^{\mathsf{M}}\,, \\ E \setminus \mathit{lin}(\mu)\,, & \theta \in \Xi_E^{\mathsf{M},\cap} \setminus \mathit{lin}(\mu),\ E \in \boldsymbol{E}\,. \end{cases}$$

Proof. By Theorem 3.2, $\theta \in \mathbb{R}^d$ is in the range if and only if (10) holds for this θ and some $a \in \mathbb{R}^d$. Since $\mathsf{m}(Q_\theta) \in \mathit{aff}(\mu)$ by Lemma 2.15, the conditions (10) and $a \in \mathit{aff}(\mu)$ are equivalent to the conditions

$$\theta \in \Xi_\mu^{\mathsf{M}} \text{ and } a - \mathsf{m}(Q_\theta) \in \mathsf{N}_\theta(\Xi) \cap \mathit{lin}(\mu)\,, \tag{11}$$

using that $a \in \mathit{aff}(\mu)$ implies $\Xi_{\mu,a} = \Xi_\mu$, while (10) and $a \notin \mathit{aff}(\mu)$ are equivalent to

$$\theta \in \Xi_E^{\mathsf{M},\cap} \text{ and } a - \mathsf{m}(Q_\theta) \in \mathsf{N}_\theta(\Xi) \cap [E \setminus \mathit{lin}(\mu)], \tag{12}$$

where $E = E_{\mu,a}$. Hence, the range is contained in $\Gamma = \Xi_\mu^{\mathsf{M}} \cup \bigcup_{E\in\boldsymbol{E}} \Xi_E^{\mathsf{M},\cap}$.

If $\theta \in \Xi_\mu^{\mathsf{M}}$ then the conditions (11) are trivially met by $a = \mathsf{m}(Q_\theta)$, and if $\theta \in \Xi_E^{\mathsf{M},\cap}$ for some $E \in \boldsymbol{E}$ then there exists $a \in \mathbb{R}^d$ satisfying the conditions (12), by the definition of $\Xi_E^{\mathsf{M},\cap}$. Hence, the range contains Γ. The proof of the first assertion of the lemma is completed by noting that $\Xi_E^{\mathsf{M},\cap} \setminus \Xi_\mu^{\mathsf{M}}$ is equal to $\Xi_E^{\mathsf{M},\cap} \setminus \mathit{lin}(\mu)$, due to the obvious inclusions

$$\Xi_E^{\mathsf{M},\cap} \cap \mathit{lin}(\mu) \subseteq \pi_\mu(\Xi_E^{\mathsf{M},\cap}) \subseteq \Xi_\mu^{\mathsf{M}}.$$

It remains to show that for θ in the range of θ^* just determined, the set of those $a \in \mathbb{R}^d$ that satisfy either (11), or (12) for some $E \in \boldsymbol{E}$, is equal to $\mathsf{m}(Q_\theta) + \mathsf{K}_\mu(\theta)$. If $\theta \in \Xi_\mu^{\mathsf{M}}$ then the set of all $a \in \mathbb{R}^d$ satisfying (11) is equal to $\mathsf{m}(Q_\theta) + [\mathsf{N}_\theta(\Xi) \cap \mathit{lin}(\mu)]$, and the set of all $a \in \mathbb{R}^d$ satisfying (12) for some $E \in \boldsymbol{E}$ has a similar form, with $\mathit{lin}(\mu)$ replaced by the union of $E \setminus \mathit{lin}(\mu)$ over those linear spaces $E \in \boldsymbol{E}$ for which $\theta \in \Xi_E^{\mathsf{M},\cap}$. This establishes the second assertion for the case $\theta \in \Xi_\mu^{\mathsf{M}}$. If θ is in the range but not in Ξ_μ^{M} then it belongs to one of the mutually disjoint sets $\Xi_E^{\mathsf{M},\cap} \setminus \mathit{lin}(\mu)$, $E \in \boldsymbol{E}$. Then no $a \in \mathbb{R}^d$ satisfies (11), while (12) is satisfied by exactly those a that belong to $\mathsf{m}(Q_\theta) + \mathsf{K}_\mu(\theta)$ with $\mathsf{K}_\mu(\theta)$ equal to $\mathsf{N}_\theta(\Xi) \cap [E \setminus \mathit{lin}(\mu)]$. ☐

Corollary 3.4. *The range of the mapping $a \mapsto Q_{\theta^*(a)}$ defined on $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$ coincides with the family $\{Q_\tau \colon \tau \in \Xi_\mu^{\mathsf{M}}\}$. If $\tau \in \Xi_\mu^{\mathsf{M}}$ then the set $\{a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi) \colon Q_{\theta^*(a)} = Q_\tau\}$ equals $\mathsf{m}(Q_\tau) + \mathsf{K}_\mu^*(\tau)$ where*

$$\mathsf{K}_\mu^*(\tau) = \bigcup \{\mathsf{K}_\mu(\theta) \colon \theta \in \theta^*(\mathit{ri}(\mu) + \mathit{bar}(\Xi)) \text{ and } \pi_\mu(\theta) = \tau\}\,.$$

Proof. Since $Q_{\theta^*(a)} = Q_{\pi_\mu\theta^*(a)}$, the first assertion follows by projecting the range of θ^*, using that $\pi_\mu(\Xi_E^{\mathsf{M},\cap}) \subseteq \Xi_\mu^{\mathsf{M}}$. The second assertion follows directly from Lemma 3.3. ☐

It is easy to see that the sets $\mathsf{K}_\mu(\theta)$ and $\mathsf{K}_\mu^*(\tau)$ are cones, both contained in $\mathit{bar}(\Xi)$.

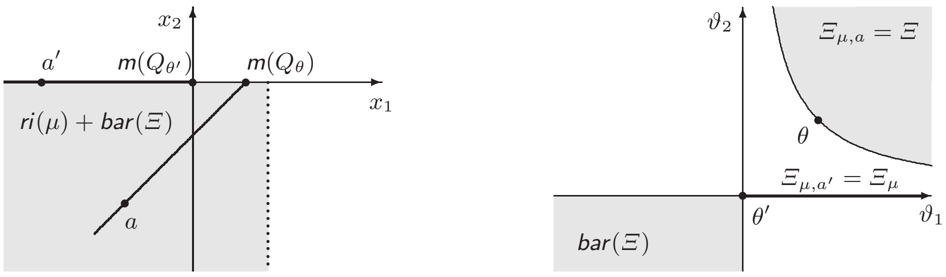

FIGURE 2.

Example 3.5. Let μ be the measure on $\mathbb{R}^2$ equal to the sum of the point masses at $(1,0)$ and $(-1,0)$. Then $aff(\mu) = lin(\mu)$ is the horizontal axis, $\Lambda(\vartheta) = \ln(e^{\vartheta_1} + e^{-\vartheta_1})$ is finite for all $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2$, and Q_ϑ has the mean $m(Q_\vartheta) = (\tanh(\vartheta_1), 0)$.

(i) Let $\Xi = \{\vartheta \colon \vartheta_1 > 0,\ \vartheta_1\vartheta_2 \geqslant 1\}$, see Fig. 2. If $a = (a_1, a_2) \in \mathbb{R}^2$ then $\Xi_{\mu,a}$ equals Ξ if $a_2 \neq 0$ and the halfaxis $\Xi_\mu = \{(\vartheta_1, 0) : \vartheta_1 \geqslant 0\}$, otherwise. Further, $ri(\mu) + bar(\Xi)$ consists of all a with $a_1 < 1$, $a_2 \leqslant 0$. It is strictly contained in its closure equal to $dom(\Psi^*)$, and strictly contains its interior. Each $a \in ri(\mu) + bar(\Xi)$ with $a_2 < 0$ can be represented as $a = (\tanh(t) - r, -rt^2)$ with $t, r > 0$, and then for $\theta = (t, t^{-1}) \in \Xi_{\mu,a}$ the vector $a - m(Q_\theta) = -r(1, t^2)$ is normal to Ξ at θ, thus $\theta^*(a) = \theta$ by Theorem 3.2. If $a = (a_1, 0)$, $0 < a_1 < 1$, then $a = m(Q_\theta)$ for $\theta = (t, 0) \in \Xi_\mu$ with suitable $t > 0$, and *(ii)* of Theorem 3.2 trivially holds, thus $\theta^*(a) = \theta$. Finally, if $a' = (a_1, 0)$, $a_1 \leqslant 0$, then $\theta' = (0,0)$ with $m(Q_{\theta'}) = (0,0)$ satisfies *(ii)* of Theorem 3.2, hence $\theta^*(a') = \theta'$. Note that while θ' is in $\Xi_{\mu,a'} = \Xi_\mu$, it is not equal to the projection of any $\vartheta \in \Xi$, even though Ξ is a closed set.

(ii) Let $\Xi = \{\vartheta \colon |\vartheta_2| < \vartheta_1\}$, see Fig. 3. Then $\Xi_{\mu,a}$ equals $cl(\Xi)$ if $a_2 \neq 0$ and Ξ_μ, the halfaxis as in *(i)*, otherwise. Further, $ri(\mu) + bar(\Xi)$ consists of all $a \in \mathbb{R}^2$ satisfying $a_1 + |a_2| < 1$. This is an open set, and $dom(\Psi^*)$ equals its closure. If $a_1 + |a_2| \leqslant 0$ then for $\theta' = (0,0)$ the vector $a - m(Q_{\theta'}) = a$ is normal to Ξ at θ', hence $\theta^*(a) = \theta'$. If $0 < a_1 + |a_2| < 1$ then a can be represented by some $t, r > 0$ as $(\tanh(t), 0)$ or $(\tanh(t) - r, r)$ or $(\tanh(t) - r, -r)$ according to a_2 is 0, positive or negative, respectively. Then, for θ equal to $(t, 0)$, (t, t) or $(t, -t)$ the vector $a - m(Q_\theta)$ is equal to $(0,0)$, $(-r, r)$ or $(-r, -r)$, respectively. In all cases the vector is normal to Ξ at θ, and hence $\theta^*(a) = \theta$. It follows from Theorem 3.2 that the range of θ^* is union of the three halflines Ξ_μ, $\{(r, r) \colon r > 0\}$ and $\{(r, -r) \colon r > 0\}$, see also Lemma 3.3. It follows also that the set $\{a \colon Q_{\theta^*(a)} = Q_\tau\}$ is the sum of $m(Q_\tau)$ and the cone $K^*_\mu(\tau)$ equal to the union of $\{(0,0)\}$ with the two halflines $\{(-r, r) \colon r > 0\}$ and $\{(-r, -r) \colon r > 0\}$, see also Corollary 3.4. The range of θ^* and $K^*_\mu(\tau)$ are not convex.

Note that while the canonically convex family $\mathcal{E}_\Xi$ is the same in *(i)* and *(ii)*, the sets $ri(\mu) + bar(\Xi)$ where $\theta^* = \theta^*_{\mu,\Xi}$ is defined are different. Even for a in their intersection, not only $\theta^*(a)$ but also the pm $Q_{\theta^*(a)}$ does depend on the choice of the parameter set Ξ in general.

3.3. In the final part of this section, the above results are discussed and related to previous ones in the special cases $\Xi = dom(\Lambda)$ and $\Xi \subseteq lin(\mu)$, sharing the feature that the set $\Xi_{\mu,a}$ in the definition of $\theta^*(a)$ does not depend on a. These cases cover those treated

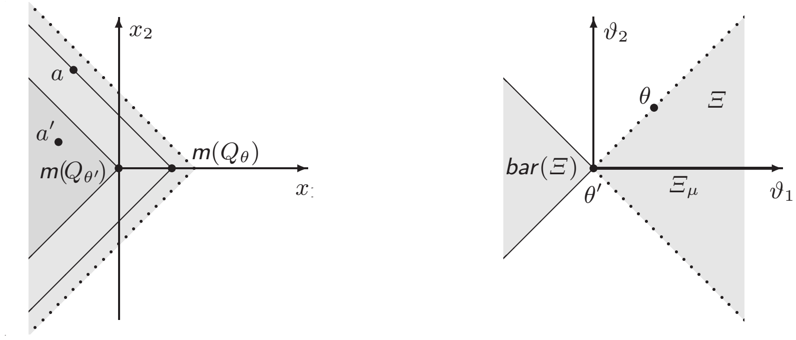

FIGURE 3.

in the literature, where $lin(\mu) = \mathbb{R}^d$ is assumed except for a few occasions addressing full families.

Consider first $\Xi = dom(\Lambda)$, the case of a full family. Then the set $ri(\mu) + bar(\Xi)$ on which the mapping θ^* is defined is equal to $ri(\mu)$. This follows from

$$(13) \qquad bar(dom(\Lambda)) \subseteq rec(dom(\Lambda^*)) \subseteq rec(ri(dom(\Lambda^*))) \quad \text{and} \quad ri(dom(\Lambda^*)) = ri(\mu)$$

where the first inclusion holds by [1, Theorem 5.19], the second one by Lemma 2.3, and for the last equality see for example [6, Proposition 1]. Further, $\Xi_{\mu,a} = dom(\Lambda) \cap lin(\mu)$ for each $a \in ri(\mu)$, thus the mapping θ^* ranges in this intersection. Using that $\tilde{\Xi}_{\mu,a} = \Xi = dom(\Lambda)$ for each $a \in \mathbb{R}^d$, Theorem 3.2 yields the result of [1, Theorem 9.30] that an MLE ϑ^*, attaining the maximum of $\langle \vartheta, a \rangle - \Lambda(\vartheta)$ subject to $\vartheta \in \mathbb{R}^d$, exists if and only if a belongs to $ri(\mu)$. By the equivalence *(i)⇔(iii)*, $\theta^*(a) + lin(\mu)^\perp$ is the set of all such MLE's. It also follows from Theorem 3.2 that the conditions

$$(14) \qquad lin(dom(\Lambda)) \subseteq M(Q_\theta) \quad \text{and} \quad a - m(Q_\theta) \in N_\theta(dom(\Lambda))$$

are necessary and sufficient for $a \in \mathbb{R}^d$ and $\theta \in dom(\Lambda)$ to satisfy $\Lambda^*(a) = \langle \theta, a \rangle - \Lambda(\theta)$. Hence, each θ in the range of θ^* satisfies the inclusion in (14). Considering $a = m(Q_\theta)$, it follows that the range of θ^* consists of all $\theta \in dom(\Lambda) \cap lin(\mu)$ that satisfy the inclusion in (14), in which case $\theta^*(m(Q_\theta)) = \theta$.

In terms of subdifferentials, recalled in Subsection 2.5,

$$(15) \quad \partial\Lambda(\theta) = \begin{cases} m(Q_\theta) + N_\theta(dom(\Lambda))\,, & \text{if } \theta \in dom(\Lambda) \text{ and } lin(dom(\Lambda)) \subseteq M(Q_\theta)\,, \\ \emptyset\,, & \text{otherwise,} \end{cases}$$

and the inverse of θ^* is the restriction of the mapping $\theta \mapsto \partial\Lambda(\theta)$ to $dom(\Lambda) \cap lin(\mu)$ in the sense

$$\{a \in ri(\mu) \colon \theta^*(a) = \theta\} = \partial\Lambda(\theta), \qquad \theta \in dom(\Lambda) \cap lin(\mu)\,.$$

When $dom(\Lambda)$ has nonempty interior, or equivalently $lin(dom(\Lambda)) = \mathbb{R}^d$, other well-known results follow. The range of θ^* consists of those $\theta \in dom(\Lambda) \cap lin(\mu)$ for which Q_θ has a mean. The subdifferential $\partial\Lambda(\theta)$ is nonempty if and only if $\theta \in dom(\Lambda)$ and Q_θ has a mean. If θ is an interior point of $dom(\Lambda)$ then $\partial\Lambda(\theta) = \{m(Q_\theta)\}$, due to $N_\theta(dom(\Lambda)) = \{0\}$, and thus Λ is differentiable at θ with its gradient equal to the mean of Q_θ, existing by Corollary 2.20. Since the sets $\partial\Lambda(\theta)$ with $\theta \in dom(\Lambda) \cap lin(\mu)$ are disjoint and cover $ri(\mu)$, each $a \in ri(\mu)$ is the mean of some pm in $\mathcal{E}_\mu$ if and only if the family $\mathcal{E}_\mu$ is steep, in the sense that no $Q_\vartheta \in \mathcal{E}$ with ϑ on the boundary of $dom(\Lambda)$ has a mean.

Consider next the case $\Xi \subseteq \mathit{lin}(\mu)$. Then, $\Xi_{\mu,a} = \Xi_\mu = \mathit{cl}(\Xi) \cap \mathit{dom}(\Lambda)$ for each $a \in \mathbb{R}^d$, and

$$\Xi_E^{M,\cap} = \{\theta \in \Xi_\mu^M \colon N_\theta(\Xi) \cap [E \setminus \mathit{lin}(\mu)] \neq \emptyset\}\,, \qquad E \in \boldsymbol{E}.$$

It follows, as a special case of Lemma 3.3, that the range of θ^* is equal to Ξ_μ^M, and for $\theta \in \Xi_\mu^M$ the set of all $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$ with $\theta^*(a) = \theta$ is equal to $m(Q_\theta) + N_\theta(\Xi)$.

In accordance to previous notations, let Ψ respectively Ψ_{Ξ_μ} denote the function equal to Λ on the set Ξ respectively Ξ_μ, and to $+\infty$ elsewhere. Then, for $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$, a maximizer of $\vartheta \mapsto \langle\vartheta, a\rangle - \Lambda(\vartheta)$ over Ξ is a parameter $\theta \in \mathbb{R}^d$ satisfying $\Psi^*(a) = \langle\theta, a\rangle - \Psi(\theta)$, while $\theta^*(a) = \theta$ is equivalent to $\Psi^*(a) = \langle\theta, a\rangle - \Psi_{\Xi_\mu}(\theta)$. These conditions are, in turn, equivalent to $a \in \partial\Psi(\theta)$ respectively $a \in \partial\Psi_{\Xi_\mu}(\theta)$, see Subsection 2.5. In particular, the subdifferentials coincide if $\theta \in \Xi$ while $\partial\Psi(\theta)$ is empty otherwise. By the special case of Lemma 3.3 mentioned above,

$$\partial\Psi_{\Xi_\mu}(\theta) = \begin{cases} m(Q_\theta) + N_\theta(\Xi)\,, & \theta \in \Xi_\mu^M\,, \\ \emptyset\,, & \text{otherwise,} \end{cases} \tag{16}$$

which implies a similar representation of $\partial\Psi(\theta)$, with Ξ_μ^M replaced by the set Ξ^M of those $\theta \in \Xi$ that satisfy $\mathit{lin}(\Xi) \subseteq M(Q_\theta)$. In the nondegenerate case, when Ξ intersects the interior of $\mathit{dom}(\Lambda)$, the last inclusion is equivalent to existence of the mean of the pm Q_θ by Corollary 2.20.

The above considerations have also the following consequences. The set $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$ partitions into the nonempty subdifferentials $\partial\Psi_{\Xi_\mu}(\theta) = m(Q_\theta) + N_\theta(\Xi)$, $\theta \in \Xi_\mu^M$, where in the nondegenerate case the last condition means that the pm Q_θ parameterized by $\theta \in \Xi_\mu$ has a mean. Further, an MLE exists if and only if a belongs to the subdifferential $\partial\Psi(\theta)$ at some $\theta \in \Xi$ which extends and strengthens the first assertion of [1, Theorem 9.18]. Finally, the condition $a \in \partial\Psi_{\Xi_\mu}(\theta)$ is equivalent to $\theta \in \partial\Psi^*_{\Xi_\mu}(a)$ by [15, Theorem 23.5], and Ψ^* is equal to $\Psi^*_{\Xi_\mu}$ by Lemma 2.13. As $\theta = \theta^*(a)$ is uniquely determined by $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$, this implies that Ψ^* is differentiable on $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$, with the gradient $\nabla\Psi^*(a) = \theta^*(a)$. This extends the third assertion of [1, Theorem 9.18] where $\mathit{lin}(\mu) = \mathbb{R}^d$ and Ξ with nonempty interior are assumed.

In the nondegenerate case, and more generally when Ξ intersects $\mathit{ri}(\mathit{dom}(\Lambda))$, the above representations of $\partial\Psi_{\Xi_\mu}(\theta)$ and $\partial\Psi(\theta)$ follow also directly from (15). To see this, recall that in that case for $\theta \in \Xi$ Lemma 2.22 gives that $\partial\Psi(\theta) = \partial\Lambda(\theta) + N_\theta(\Xi)$, and Lemma 2.19 gives that the inclusion in (15) holds if and only if $\mathit{lin}(\Xi) \subseteq M(Q_\theta)$ thus $\theta \in \Xi^M$.

4. The GMLE in general

By Fact 2.7, the restriction of μ to the closure of a face F of $\mathit{cc}(\mu)$ is a nonzero measure whose convex core is equal to F. For this restriction $\nu = \mu^{\mathit{cl}(F)}$, with $\mathit{ri}(\nu)$, $\mathit{aff}(\nu)$, $\mathit{lin}(\nu)$ equal to $\mathit{ri}(F)$, $\mathit{aff}(F)$, $\mathit{lin}(F)$, in the sequel the convenient notations Λ_F, $Q_{F,\vartheta}$, $\Psi^*_{F,\Xi}$, $\theta^*_{F,\Xi}$, etc. are used instead of Λ_ν, $Q_{\nu,\vartheta}$, $\Psi^*_{\nu,\Xi}$, $\theta^*_{\nu,\Xi}$, etc.

When $a \in \mathit{cc}(\mu) + \mathit{bar}(\Xi)$, among the faces G of $\mathit{cc}(\mu)$ satisfying $a \in \mathit{ri}(G) + \mathit{bar}(\Xi)$ there exists the inclusion-largest one, by Lemma 2.1. This face is denoted by $G^*(a) = G^*_{\mu,\Xi}(a)$.

Theorem 4.1. *If $a \in \mathit{dom}(\Psi^*)$ then $a \in \mathit{cc}(\mu) + \mathit{bar}(\Xi)$, the face $G = G^*(a)$ of $\mathit{cc}(\mu)$ is Ξ-accessible, $\Psi^*_{G,\Xi}(a)$ is equal to $\Psi^*(a) = \Psi^*_{\mu,\Xi}(a)$, and with $\theta = \theta^*_{G,\Xi}(a)$*

$$\Psi^*(a) - \big[\langle\vartheta, a\rangle - \Lambda(\vartheta)\big] \geqslant D(Q_{G,\theta} \| Q_\vartheta)\,, \qquad \vartheta \in \Xi\,. \tag{17}$$

This proves the existence of GMLE by explicitly identifying it. Indeed, comparison of (2) and (17) gives, referring to Remark 1.2 for uniqueness, that $Q_{G,\theta} = R^*(a)$.

Corollary 4.2. *The* GMLE $R^*_{\mu,\Xi}(a)$ *belongs to* $\mathcal{E} = \mathcal{E}_\mu$ *if and only if* $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$.

In particular, for a full exponential family $\mathcal{E}$ it follows that $R^*(a) \in \mathcal{E}$ if and only if $a \in \mathsf{ri}(\mu)$, or equivalently, if and only if an MLE exists, see the passage containing eq. (13) in Subsection 3.3.

Remark 4.3. The inequality (17) holds even for $\vartheta \in \tilde{\Xi}_{\mu,a}$. In fact, by Lemma 2.11, the bracket and the pm Q_ϑ do not change when $\vartheta \in \Xi$ is replaced by $\pi_{\mu,a}(\vartheta)$, thus (17) holds for ϑ in $\pi_{\mu,a}(\Xi)$. By limiting along segments, continuity of Λ along segments and lower semicontituity of I-divergence, (17) holds for ϑ in $\Xi_{\mu,a}$ and, in turn, for $\vartheta \in \tilde{\Xi}_{\mu,a}$ by Lemma 2.11. A further immediate extension of (17) is

$$\Psi^*(a) - \big[\langle\vartheta, a\rangle - \Lambda_F(\vartheta)\big] \geqslant D(Q_{G,\theta} \| Q_{F,\vartheta}) \,, \qquad \vartheta \in \tilde{\Xi}_{F,a} \,, \tag{18}$$

where F is any face of $\mathsf{cc}(\mu)$ containing $G = G^*(a)$. Indeed, the obvious identity

$$D(P\|Q_{F,\vartheta}) - \Lambda_F(\vartheta) = D(P\|Q_\vartheta) - \Lambda(\vartheta) \,,$$

valid for any face F of $\mathsf{cc}(\mu)$, pm $P \ll \mu^{\mathsf{cl}(F)}$, and $\vartheta \in \mathsf{dom}(\Lambda)$, implies with $P = Q_{G,\theta}$ that for faces $F \supseteq G$ of $\mathsf{cc}(\mu)$, (18) with $\vartheta \in \Xi$ is equivalent to (17). Then, the extension to $\vartheta \in \tilde{\Xi}_{F,a}$ follows by limiting as above.

Theorem 4.1 will be proved via induction on the dimension of $\mathsf{aff}(\mu)$. Five lemmas are sent forth, the first three of elementary geometric content.

Lemma 4.4. *If* $B \subseteq \mathsf{aff}(\mu)$ *then* $a \in B + \mathsf{bar}(\Xi)$ *is equivalent to* $a \in B + \mathsf{bar}(\pi_{\mu,a}(\Xi))$.

Proof. If $x \in E_{\mu,a}$ then $\langle\vartheta, x\rangle = \langle\pi_{\mu,a}(\vartheta), x\rangle$ for all $\vartheta \in \mathbb{R}^d$, hence $x \in \mathsf{bar}(\Xi)$ takes place if an only if $x \in \mathsf{bar}(\pi_{\mu,a}(\Xi))$. Applying this to $x = a - b$ with $b \in B$, when $x \in E_{\mu,a}$ due to $B \subseteq \mathsf{aff}(\mu)$, the assertion follows. □

Lemma 4.5. *For a nonempty convex set* C *and cone* K *containing* 0, *a halfspace* $H_\leqslant = \{x : \langle\theta, x - a\rangle \leqslant 0\}$ *with* $\theta \neq 0$ *contains* $C + K$ *if and only if* $C \subseteq H_\leqslant$ *and* $\theta \in \mathsf{N}_0(K)$.

Proof. If $C + K \subseteq H_\leqslant$ then $C \subseteq H_\leqslant$ as $0 \in K$. In addition, for any c from the nonempty set C, the inequality $\langle\theta, c + y - a\rangle \leqslant 0$ holds for all $y \in K$. Since K is a cone, it follows that $\langle\theta, y\rangle \leqslant 0$, $y \in K$, and thus $\theta \in \mathsf{N}_0(K)$. The converse implication is obvious. □

Lemma 4.6. *For* C, K *and* $H_\leqslant$ *as in Lemma 4.5, if* $C+K \subseteq H_\leqslant$ *and* $\mathsf{ri}(C)+K$ *intersects the boundary hyperplane* H *of* $H_\leqslant$ *then* $C \subseteq H$.

Proof. Suppose some $x \in \mathsf{ri}(C) + K$, say $x = c + y$ where $c \in \mathsf{ri}(C)$ and $y \in K$, belongs to H, thus $\langle\theta, c + y - a\rangle = 0$. By Lemma 4.5, $C \subseteq H_\leqslant$ and $\theta \in \mathsf{N}_0(K)$, thus $\langle\theta, c - a\rangle \leqslant 0$ and $\langle\theta, y\rangle \leqslant 0$. Hence $\langle\theta, c - a\rangle = 0$, that is, H contains $c \in \mathsf{ri}(C)$. This and $C \subseteq H_\leqslant$ imply the assertion. □

Lemma 4.7. *If* $a \in \mathsf{dom}(\Psi^*_{\mu,\Gamma})$ *for a convex subset* Γ *of* $\mathsf{dom}(\Lambda)$, *and* $\mathsf{cs}(\mu)$ *is contained in a halfspace* $H_\leqslant = \{x\colon \langle\theta, x - a\rangle \leqslant 0\}$ *with a nonzero* $\theta \in \mathsf{rec}(\mathsf{ri}(\Gamma))$, *then the boundary hyperplane* H *of* $H_\leqslant$ *intersects* $\mathsf{cc}(\mu)$ *in a face* F *of* $\mathsf{cc}(\mu)$ *such that* $\Psi^*_{\mu,\Gamma}(a) = \Psi^*_{F,\Gamma}(a)$.

Proof. If $\vartheta \in \mathit{ri}(\Gamma)$ and $t \geqslant 0$ then $\vartheta + t\theta \in \mathit{ri}(\Gamma)$ due to $\theta \in \mathit{rec}(\mathit{ri}(\Gamma))$, thus

$$\Psi^*_{\mu,\Gamma}(a) \geqslant \langle \vartheta + t\theta, a\rangle - \Lambda(\vartheta + t\theta) = -\ln \int_{\mathbb{R}^d} e^{\langle \vartheta, x-a\rangle + t\langle \theta, x-a\rangle}\,\mu(dx)\,.$$

Since $H_{\leqslant}$ contains $\mathit{cs}(\mu)$, it has full μ-measure. The integrand is bounded on $H_{\leqslant}$ by $e^{\langle \vartheta, x-a\rangle}$ which is μ-integrable due to $\vartheta \in \Gamma \subseteq \mathit{dom}(\Lambda)$. By dominated convergence, the integral converges to $\int_H e^{\langle \vartheta, x-a\rangle}\,\mu(dx)$ when $t \to \infty$. It follows that

$$\Psi^*_{\mu,\Gamma}(a) \geqslant -\ln \int_H e^{\langle \vartheta, x-a\rangle}\,\mu(dx), \qquad \vartheta \in \mathit{ri}(\Gamma)\,,$$

and as $\Psi^*_{\mu,\Gamma}(a)$ is finite by assumption, this implies $\mu(H) > 0$. Hence, H is a supporting hyperplane of $\mathit{cs}(\mu)$. By Fact 2.9, $F = H \cap \mathit{cc}(\mu)$ is a face of $\mathit{cc}(\mu)$ and $\mu(H \setminus \mathit{cl}(F)) = 0$. Therefore, the above inequality rewrites to

$$\Psi^*_{\mu,\Gamma}(a) \geqslant -\ln \int_{\mathit{cl}(F)} e^{\langle \vartheta, x-a\rangle}\,\mu(dx) = \langle \vartheta, a\rangle - \Lambda_F(\vartheta), \qquad \vartheta \in \mathit{ri}(\Gamma)\,.$$

This and Lemma 2.13 imply that $\Psi^*_{\mu,\Gamma}(a) \geqslant \Psi^*_{F,\mathit{ri}(\Gamma)}(a) = \Psi^*_{F,\Gamma}(a)$, while $\Lambda \geqslant \Lambda_F$ implies that $\Psi^*_{\mu,\Gamma}(a) \leqslant \Psi^*_{F,\Gamma}(a)$. □

Lemma 4.8. *If a belongs to $\mathit{dom}(\Psi^*_{\mu,\Xi})$ but not to $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$ then some unit vector $\tau \in \mathit{rec}(\pi_\mu(\mathit{ri}(\Xi)))$ exposes a proper face F of $\mathit{cc}(\mu)$ such that $\Psi^*_{\mu,\Xi}(a) = \Psi^*_{F,\Xi}(a)$ and F contains each face G of $\mathit{cc}(\mu)$ with $a \in \mathit{ri}(G) + \mathit{bar}(\Xi)$.*

Proof. Let Γ denote the convex subset $\pi_{\mu,a}(\Xi)$ of $\mathit{dom}(\Lambda)$. By Corollary 2.12, $\Psi^*_{\mu,\Xi}(a)$ equals $\Psi^*_{\mu,\Gamma}(a)$. This and Lemma 4.4 with $B = \mathit{ri}(\mu)$ imply that the assumptions equivalently mean that a is in $\mathit{dom}(\Psi^*_{\mu,\Gamma})$ but not in $\mathit{ri}(\mu) + \mathit{bar}(\Gamma)$. On account of the latter, there exists a halfspace $H_{\leqslant} = \{x \in \mathbb{R}^d\colon \langle \theta, x-a\rangle \leqslant 0\}$ with nonzero θ that contains $\mathit{ri}(\mu) + \mathit{bar}(\Gamma)$. By Lemma 4.5 with $C = \mathit{ri}(\mu)$ and $K = \mathit{bar}(\Gamma)$, the halfspace $H_{\leqslant}$ contains $\mathit{ri}(\mu)$ and hence also $\mathit{cs}(\mu)$, and θ belongs to $N_0(\mathit{bar}(\Gamma))$ that by Lemma 2.3 is equal to $\mathit{rec}(\mathit{ri}(\Gamma))$. Then, Lemma 4.7 implies that the intersection of $\mathit{cc}(\mu)$ with the boundary hyperplane H of $H_{\leqslant}$ is a face F of $\mathit{cc}(\mu)$, and $\Psi^*_{\mu,\Gamma}(a) = \Psi^*_{F,\Gamma}(a)$.

As Γ is a subset of $E_{\mu,a}$, the linear span of $\{x - a\colon x \in \mathit{ri}(\mu)\}$, so is also the recession cone $\mathit{rec}(\mathit{ri}(\Gamma))$, thus the nonzero vector θ contained in that cone satisfies $\langle \theta, x-a\rangle \neq 0$ for some $x \in \mathit{ri}(\mu)$. The latter implies $\langle \theta, x-y\rangle \neq 0$ for all $y \in H$, in particular, for y in $F = H \cap \mathit{cc}(\mu) \neq \emptyset$. Thus θ is not orthogonal to $\mathit{lin}(\mu)$. Using that $\mathit{cc}(\mu) \subseteq H_{\leqslant}$, it follows that F is a proper face of $\mathit{cc}(\mu)$, exposed by a unit vector proportional to θ, and hence also by the unit vector τ proportional to $\pi_\mu(\theta)$. This vector satisfies

$$\tau \in \pi_\mu(\mathit{rec}(\mathit{ri}(\Gamma))) \subseteq \mathit{rec}(\pi_\mu(\mathit{ri}(\Gamma))) = \mathit{rec}(\pi_\mu(\mathit{ri}(\Xi))),$$

where the inclusion holds by Lemma 2.2 and the equality by obvious interchange of projections and relative interiors.

The equality $\Psi^*_{\mu,\Xi}(a) = \Psi^*_{F,\Xi}(a)$ is a consequence of the two equalities in the first passage of the proof and of $\Psi^*_{F,\Gamma}(a) = \Psi^*_{F,\Xi}(a)$, obtained from Corollary 2.12 with $\mu^{\mathit{cl}(F)}$ in the role of μ.

For the last assertion, consider a face G of $\mathit{cc}(\mu)$ with $a \in \mathit{ri}(G) + \mathit{bar}(\Xi)$, equivalent to $a \in \mathit{ri}(G) + \mathit{bar}(\Gamma)$ by Lemma 4.4 with $B = \mathit{ri}(G)$. Since $G + \mathit{bar}(\Gamma)$ is contained in $\mathit{cc}(\mu) + \mathit{bar}(\Gamma) \subseteq H_{\leqslant}$ and $\mathit{ri}(G) + \mathit{bar}(\Gamma)$ contains $a \in H$, Lemma 4.6 with $C = G$ and $K = \mathit{bar}(\Gamma)$ implies $G \subseteq H$. This proves that G is a subset of F. □

Proof of Theorem 4.1. If $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ then obviously $a \in \mathsf{cc}(\mu) + \mathsf{bar}(\Xi)$ and the face $G^*_{\mu,\Xi}(a)$ coincides with $\mathsf{cc}(\mu)$ and is Ξ-accessible. Since a and $\theta^*_{\mu,\Xi}(a) = \theta$ satisfy *(iii)* of Theorem 3.2 they also satisfy *(i)* and *(iv)* which immediately rewrites to

$$\Psi^*(a) - \big[\langle \vartheta, a\rangle - \Lambda(\vartheta)\big] \geqslant D(Q_\theta \| Q_\vartheta)\,, \qquad \vartheta \in \Xi\,,$$

proving (17). In particular, the assertion of Theorem 4.1 holds when the dimension of $\mathsf{aff}(\mu)$ is zero, since then $\mathsf{dom}(\Psi^*_{\mu,\Xi})$ trivially equals $\mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$. Induction argument on the dimension is applied, assuming validity of the assertion for any measure ν with $\mathsf{aff}(\nu)$ of smaller dimension than $\mathsf{aff}(\mu)$.

In the induction step, it suffices to consider the case when a is in $\mathsf{dom}(\Psi^*)$ but not in $\mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$. Then Lemma 4.8 implies existence of a proper face F of $\mathsf{cc}(\mu)$ exposed by some unit vector $\tau \in \mathsf{rec}(\pi_\mu(\mathsf{ri}(\Xi)))$ such that $\Psi^*(a) = \Psi^*_{\mu,\Xi}(a)$ is equal to $\Psi^*_{F,\Xi}(a)$. In particular, a belongs to $\mathsf{dom}(\Psi^*_{F,\Xi})$, thus the induction hypothesis applies to $\nu = \mu^{\mathsf{cl}(F)}$. Hence, $a \in F + \mathsf{bar}(\Xi)$, the largest face G of F satisfying $a \in \mathsf{ri}(G) + \mathsf{bar}(\Xi)$ is Ξ-accessible, $\Psi^*_{F,\Xi}(a) = \Psi^*_{G,\Xi}(a)$, and with $\theta = \theta^*_{G,\Xi}(a)$

$$\Psi^*_{F,\Xi}(a) - \big[\langle \vartheta, a\rangle - \Lambda_F(\vartheta)\big] \geqslant D(Q_{G,\theta} \| Q_{F,\vartheta})\,, \qquad \vartheta \in \Xi\,.$$

As a belongs to $F + \mathsf{bar}(\Xi)$ it belongs $\mathsf{cc}(\mu) + \mathsf{bar}(\Xi)$, proving the first assertion. Then, the face $G^*(a)$ of $\mathsf{cc}(\mu)$ is well defined. By the last assertion of Lemma 4.8, this face is contained in F whence its maximality implies $G^*(a) = G$. The Ξ-accessibility of G from $\mathsf{cc}(\mu)$ is a consequence of that of G from F and $\tau \in \mathsf{rec}(\pi_\mu(\mathsf{ri}(\Xi)))$, proving the second assertion. The equality $\Psi^*_{G,\Xi}(a) = \Psi^*(a)$ follows by combining two above equalities of the same type. Finally, since $\Psi^*_{F,\Xi}(a) = \Psi^*(a)$, the last display means that (18) holds for $\vartheta \in \Xi$. As shown in Remark 4.3, this is equivalent to (17). □

Theorem 4.9. $\mathsf{dom}(\Psi^*) = \mathsf{cc}(\mu) + \mathsf{bar}(\Xi) = \bigcup \mathsf{ri}(G) + \mathsf{bar}(\Xi)$ *where the union runs over the Ξ-accessible faces G of $\mathsf{cc}(\mu)$.*

Proof. The union is obviously contained in $\mathsf{cc}(\mu) + \mathsf{bar}(\Xi)$ and contains $\mathsf{dom}(\Psi^*)$ by Theorem 4.1. Thus, it suffices to prove that $\mathsf{cc}(\mu) + \mathsf{bar}(\Xi)$ is a subset of $\mathsf{dom}(\Psi^*)$. This follows from the inclusions $\mathsf{cc}(\mu) \subseteq \mathsf{dom}(\Psi^*)$ and $\mathsf{bar}(\Xi) \subseteq \mathsf{rec}(\mathsf{dom}(\Psi^*))$. The former holds since $\mathsf{cc}(\mu) \subseteq \mathsf{dom}(\Lambda^*)$ [6, Proposition 1*(i)*] and the latter is a consequence of $\Xi = \mathsf{dom}(\Psi)$, see Remark 1.1, and $\mathsf{bar}(\mathsf{dom}(\Psi)) \subseteq \mathsf{rec}(\mathsf{dom}(\Psi^*))$ [1, Theorem 5.19]. □

Remark 4.10. In the special case of a full exponential family, Theorem 4.9 implies that $\mathsf{dom}(\Lambda^*)$ is equal to $\mathsf{cc}(\mu) + \mathsf{bar}(\mathsf{dom}(\Lambda))$. Once this special case is known, the first equality of Theorem 4.9 is equivalent to $\mathsf{dom}(\Psi^*) = \mathsf{dom}(\Lambda^*) + \mathsf{bar}(\Xi)$, due to the obvious inclusion $\mathsf{bar}(\mathsf{dom}(\Lambda)) \subseteq \mathsf{bar}(\Xi)$. In particular, the latter equality does not require the condition under which it has been proven in Lemma 2.22, let alone the condition under which it has been known previously, see Remark 2.23.

Proposition 4.11. *A set $A \subseteq \mathbb{R}^d$ is equal to $\mathsf{dom}(\Lambda^*_\mu)$ for some nonzero Borel measure μ with $\mathsf{dom}(\Lambda_\mu) \neq \emptyset$ if and only if $A = C + K$ where $C \subseteq \mathbb{R}^d$ is a nonempty convex set with at most countably many faces and K is a nonempty closed convex cone contained in $\mathsf{rec}(\mathsf{cl}(C))$.*

For example, closed balls cannot be written in this form if $d \geqslant 2$.

Proof. For a measure μ with the above property the cone $K = \mathit{bar}(\mathit{dom}(\Lambda_\mu))$ is a subset of $\mathit{rec}(\mathit{dom}(\Lambda^*_\mu))$ by (13) which is contained in $\mathit{rec}(\mathit{cl}(\mathit{dom}(\Lambda^*_\mu)))$ by Lemma 2.3. Here, $\mathit{cl}(\mathit{dom}(\Lambda^*_\mu))$ equals $\mathit{cs}(\mu)$ by [6, Proposition 1] which is the closure of $C = \mathit{cc}(\mu)$ by Fact 2.6. Hence, $K \subseteq \mathit{rec}(\mathit{cl}(C))$. By [4, Theorem 1], C has at most a countable number of faces. Theorem 4.9 gives the equality $\mathit{dom}(\Lambda^*_\mu) = C + K$ and one implication follows.

In the opposite direction, assuming nonempty C and K have the above properties, an inspection of the proof of [4, Theorem 1] shows that there exists a Borel pm ν such that $\mathit{cc}(\nu) = C$ and $\mathit{dom}(\Lambda_\nu) = \mathbb{R}^d$. If $K = \{0\}$ then, by Theorem 4.9, $\mathit{dom}(\Lambda^*_\mu) = C+K$ where $\mu = \nu$. Otherwise, a sequence of unit vectors $b_n \in \mathit{rec}(\mathit{cl}(C))$ exists such that K is the inclusion-smallest closed convex cone containing them. Let $a \in \mathit{ri}(C)$ and ν_n be the pm sitting on the halfline $\{a+tb_n : t \geqslant 0\}$ with a density $\frac{dt}{(t+1)^2}$. Then $\mathit{dom}(\Lambda_{\nu_n})$ is the halfspace given by $\{\vartheta \in \mathbb{R}^d : \langle \vartheta, b_n\rangle \leqslant 0\}$. Let μ equal to $\nu + \sum 2^{-n}\nu_n$. Then, $\mathit{dom}(\Lambda_\mu)$ equals the intersection of the halfspaces and, in turn, $\mathit{bar}(\mathit{dom}(\Lambda_\mu))$ is the cone K by [15, Section 14]. Since the halflines are contained in $\mathit{ri}(\nu)$, using that $\mathit{rec}(\mathit{cl}(C)) = \mathit{rec}(\mathit{ri}(C))$ by Lemma 2.3, the convex core of μ is C. Theorem 4.9 implies that again $\mathit{dom}(\Lambda^*_\mu) = C + K$. □

5. Properties of GMLE

5.1. Recall that notations involving the underlying measure in an index are conveniently shortened when that measure is the restriction of μ to the closure of a face of $\mathit{cc}(\mu)$, replacing the restriction by the face. For example, Ξ_G is a shorthand for $\Xi_{\mu^{\mathit{cl}(G)}}$, thus denotes $\mathit{cl}(\pi_G(\Xi)) \cap \mathit{dom}(\Lambda_G)$, the set Ξ^{M}_G consists of those $\theta \in \Xi_G$ that satisfy the inclusion $\mathit{lin}(\Xi) \subseteq \mathsf{M}(Q_{G,\theta})$, and $\mathsf{K}^*_G(\tau)$ is given as in Corollary 3.4 with μ replaced by $\mu^{\mathit{cl}(G)}$.

By [8, Theorem 2], the variation closure $\mathit{cl}_v(\mathcal{E}_\Xi)$ of any canonically convex exponential family $\mathcal{E}_\Xi = \{Q_\vartheta : \vartheta \in \Xi\}$ is equal to the union of the families $\mathcal{E}_{G,\Xi_G}$ over all Ξ-accessible faces G of $\mathit{cc}(\mu)$. While there $\Xi \subseteq \mathit{lin}(\mu)$ is assumed, the result holds with general $\Xi \subseteq \mathit{dom}(\Lambda)$ since the set Ξ_G depends on Ξ only through $\pi_\mu(\Xi)$.

Theorem 5.1. *The range of the* GMLE *mapping R^* is equal to the set of pm's $P \in \mathit{cl}_v(\mathcal{E}_\Xi)$ with $\mathit{lin}(\Xi) \subseteq \mathsf{M}(P)$, thus the union of the families $\mathcal{E}_{G,\Xi^{\mathsf{M}}_G}$ over all Ξ-accessible faces G of $\mathit{cc}(\mu)$. For $P = Q_{G,\tau}$ in this range, where G is a Ξ-accessible face of $\mathit{cc}(\mu)$ and $\tau \in \Xi^{\mathsf{M}}_G$,*

$$\{a : R^*(a) = P\} = \big[\mathsf{m}(P) + \mathsf{K}^*_G(\tau)\big] \setminus \bigcup\, [\mathit{ri}(F) + \mathit{bar}(\Xi)],$$

where the union is over those Ξ-accessible faces F of $\mathit{cc}(\mu)$ that properly contain G. This set is a shifted cone with the apex $\mathsf{m}(P)$.

As a consequence, $\mathit{dom}(\Psi^*)$ partitions into shifted cones with apices $\mathsf{m}(P)$, perhaps reducing to singletons $\{\mathsf{m}(P)\}$, extending the result of Corollary 3.4 about such partitioning of the subset $\mathit{ri}(\mu) + \mathit{bar}(\Xi)$ of $\mathit{dom}(\Psi^*)$.

Corollary 5.2. *In the nondegenerate case, the range of the* GMLE *mapping consists of those pm's from $\mathit{cl}_v(\mathcal{E}_\Xi)$ that have means.*

Proof. If Ξ intersects the interior of $\mathit{dom}(\Lambda)$ then it intersects also the interior of $\mathit{dom}(\Lambda_G)$ for each Ξ-accessible face G of $\mathit{cc}(\mu)$. Corollary 2.20 applied to restrictions of μ implies the equivalence of the inclusion $\mathit{lin}(\Xi) \subseteq \mathsf{M}(P)$ to the existence of the mean of P, and thus the assertion follows from Theorem 5.1. □

The proof of Theorem 5.1 is preceded by two lemmas.

Lemma 5.3. *If G is a proper Ξ-accessible face of $\mathsf{cc}(\mu)$ and a pm P is concentrated on $\mathsf{aff}(G)$ then $\mathsf{m}(P) \notin \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$.*

Proof. By the first assumption, some unit vector $\theta \in \mathsf{rec}(\pi_\mu(\mathsf{ri}(\Xi)))$ exposes a proper face F of $\mathsf{cc}(\mu)$ such that $G \subseteq F$. Let $H_\leqslant$ denote the halfspace $\{x\colon \langle \theta, x-a\rangle \leqslant 0\}$ that contains $\mathsf{cc}(\mu)$ and whose boundary hyperplane H intersects $\mathsf{cc}(\mu)$ in F. By Lemma 2.15, $\mathsf{m}(P) \in \mathsf{aff}(P)$, where $\mathsf{aff}(P) \subseteq \mathsf{aff}(G)$ due to the second assumption. Hence, $\mathsf{m}(P)$ belongs to $\mathsf{aff}(F)$, and thus to H. On the other hand, let $C = \mathsf{cc}(\mu)$ and $K = \mathsf{bar}(\pi_\mu(\Xi))$. By Lemma 2.3, $\mathsf{rec}(\mathsf{ri}(\pi_\mu(\Xi)))$ equals $\mathsf{N}_0(K)$, whence θ belongs to $\mathsf{N}_0(K)$. This and $C \subseteq H_\leqslant$ imply $C + K \subseteq H_\leqslant$ on account of Lemma 4.5. Since $C \nsubseteq H$, it follows from Lemma 4.6 that $\mathsf{ri}(C) + K$ does not intersect H. Therefore, $\mathsf{m}(P) \notin \mathsf{ri}(\mu) + \mathsf{bar}(\pi_\mu(\Xi))$. The proof is completed applying Lemma 4.4 to $B = \mathsf{ri}(\mu)$ and $a = \mathsf{m}(P)$, as π_μ and $\pi_{\mu,a}$ are identical projections due to $a \in \mathsf{aff}(\mu)$. ☐

Lemma 5.4. *If $P \in \mathsf{cl}_v(\mathcal{E}_\Xi)$ satisfies $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(P)$ then $\mathsf{m}(P)$ belongs to $\mathsf{ri}(P) + \mathsf{bar}(\Xi)$, this set is contained in $\mathsf{dom}(\Psi^*)$, and $R^*(\mathsf{m}(P)) = P$.*

Proof. If $P \in \mathsf{cl}_v(\mathcal{E}_\Xi)$ then $P = Q_{G,\theta}$ for a Ξ-accessible face G of $\mathsf{cc}(\mu)$ and some $\theta \in \Xi_G$. The latter and the assumption $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(Q_{G,\theta})$ mean that θ belongs to Ξ_G^{M}. Lemma 3.3 implies that the set $\{a \in \mathsf{ri}(G) + \mathsf{bar}(\Xi)\colon \theta^*_{G,\Xi}(a) = \theta\}$ contains $\mathsf{m}(P)$. Since $\mathsf{ri}(P) = \mathsf{ri}(G)$, this proves that $\mathsf{m}(P)$ belongs to $\mathsf{ri}(P) + \mathsf{bar}(\Xi)$, a subset of $\mathsf{dom}(\Psi^*)$ due to Theorem 4.9. In addition, $R^*(\mathsf{m}(P)) = P$ follows from Theorem 4.1 if it is shown that $a = \mathsf{m}(P)$ satisfies $G^*(a) = G$, or equivalently that if a face F of $\mathsf{cc}(\mu)$ properly contains G then $a \notin \mathsf{ri}(F) + \mathsf{bar}(\Xi)$. Lemma 2.4 implies that the Ξ-accessible face G of $\mathsf{cc}(\mu)$ is a Ξ-accessible face also of F, and the assertion follows from Lemma 5.3 applied to $\nu = \mu^{\mathsf{cl}(F)}$ instead of μ, since P is concentrated on $\mathsf{cl}(F) \subseteq \mathsf{aff}(F)$ and $\mathsf{ri}(\nu) = \mathsf{ri}(F)$. ☐

Proof of Theorem 5.1. If $P = R^*(a)$ for some $a \in \mathsf{dom}(\Psi^*)$ then $P \in \mathsf{cl}_v(\mathcal{E}_\Xi)$ by Remark 1.2. On account of (2), the I-divergence $D(P\|Q_\vartheta)$ is finite for $\vartheta \in \Xi$, thus Corollary 2.17 implies $\mathsf{lin}(\Xi) \subseteq \mathsf{M}(P)$. This and Lemma 5.4 prove the first assertion, equivalent to the second one due to the representation of $\mathsf{cl}_v(\mathcal{E}_\Xi)$ cited above.

For $P = Q_{G,\tau}$ with Ξ-accessible G and $\tau \in \Xi_G^{\mathsf{M}}$, an element a of $\mathsf{dom}(\Psi^*)$ satisfies $R^*(a) = P$ if and only if $G^*(a) = G$ and $Q_{\theta^*_{G,\Xi}(a)} = P$, by Theorem 4.1. The first condition means, by the definition of $G^*(a)$, that a belongs to $\mathsf{ri}(G) + \mathsf{bar}(\Xi)$ but not to the union of $\mathsf{ri}(F) + \mathsf{bar}(\Xi)$ for the Ξ-accessible faces F of $\mathsf{cc}(\mu)$ containing G properly. The second condition is equivalent to $a \in \mathsf{m}(P) + \mathsf{K}^*_G(\tau)$, by Corollary 3.4 applied to $\mu^{\mathsf{cl}(G)}$ in the role of μ. Since $\mathsf{m}(P)$ belongs to $\mathsf{ri}(P) + \mathsf{bar}(\Xi)$ by Lemma 5.4, and $\mathsf{K}^*_G(\tau)$ is contained in $\mathsf{bar}(\Xi)$, $\mathsf{ri}(G) + \mathsf{bar}(\Xi)$ contains $\mathsf{m}(P) + \mathsf{K}^*_G(\tau)$, and the claimed representation of the set $\{a\colon R^*(a) = P\}$ follows.

This set contains $\mathsf{m}(P)$ by Lemma 5.4, hence for the last assertion it suffices to prove that if a is in the intersection of $\mathsf{m}(P) + \mathsf{K}^*_G(\tau)$ and $\mathsf{ri}(F) + \mathsf{bar}(\Xi)$ for a Ξ-accessible face $F \supset G$ of $\mathsf{cc}(\mu)$ then $b_t = \mathsf{m}(P) + t[a - \mathsf{m}(P)]$ belongs to that intersection for all $t > 0$. Obviously, $b_t \in \mathsf{m}(P) + \mathsf{K}^*_G(\tau)$ since $\mathsf{K}^*_G(\tau)$ is a cone. Writing $a = c + d$ with some $c \in \mathsf{ri}(F)$ and $d \in \mathsf{bar}(\Xi)$, for $0 < r < \min\{t, 1\}$

$$b_t = [rc + (1-r)\mathsf{m}(P)] + rd + (t-r)(a - \mathsf{m}(P))\,.$$

Since $\mathsf{m}(P) \in \mathsf{ri}(P) + \mathsf{bar}(\Xi) \subseteq F + \mathsf{bar}(\Xi)$, the above bracket belongs to $\mathsf{ri}(F) + \mathsf{bar}(\Xi)$. This, $d \in \mathsf{bar}(\Xi)$ and $a - \mathsf{m}(P) \in \mathsf{K}^*_G(\tau) \subseteq \mathsf{bar}(\Xi)$ imply $b_t \in \mathsf{ri}(F) + \mathsf{bar}(\Xi)$. ☐

Example 5.5. Let us consider the situation of Example 1.4 with its notations. There, it was shown by elementary means that for $P = \nu$ the set $\{a\colon R^*(a) = P\}$ equals ∂C, which is a nonconvex cone. This equality can be simply derived also from Theorem 4.1. In fact, the convex core of μ consists of the origin c and the interior of C [4, Example 1]. Its only proper face is $G = \{c\}$. Any $a \in \partial C$ belongs to $\mathsf{dom}(\Lambda^*) = C$ and since it is not in $\mathsf{ri}(\mu) + \mathsf{bar}(\Xi) = \mathsf{ri}(\mu)$ it follows that $G^*(a) = G$. Theorem 4.1 implies $R^*(a) = P$. But $\{a\colon R^*(a) = P\}$ is contained in ∂C by Corollary 4.2.

For an illustration, the equality $\{a\colon R^*(a) = P\} = \partial C$ is now derived directly from Theorem 5.1. The point mass $P = \nu$ belongs to the range of R^* because $P = Q_{G,\tau}$ for $\tau = (0,0,0)$ where the face G of $\mathsf{cc}(\mu)$ is Ξ-accessible and $\tau \in \Xi^{\mathsf{M}}_G$, which is a singleton. By Theorem 5.1,

$$\{a\colon R^*(a) = P\} = [c + \mathsf{K}^*_G(\tau)] \setminus [\mathsf{ri}(C) + \mathsf{bar}(\Xi)] = \mathsf{K}^*_G(\tau) \setminus \mathsf{ri}(C)$$

and it suffices to show that $\mathsf{K}^*_G(\tau) = C$. To this end, observe that the mapping $\theta^*_{G,\Xi}$ is defined on $\mathsf{ri}(G) + \mathsf{bar}(\Xi) = C$. Obviously, $\theta^*_{G,\Xi}(c) = \tau$. For $b \in C \setminus G$, $\Xi_{G,b} = \mathsf{cl}(\pi_{G,b}(\Xi))$ is the orthogonal projection of $(1,0,0) - C$ to $E_{G,b}$, the line that passes through b and c, and it is easy to see directly from the definition of $\theta^*_{G,\Xi}$ that $\theta^*_{G,\Xi}(b)$ is the projection of $(1,0,0)$ to this line. Therefore, the range of $\theta^*_{G,\Xi}$ is the union of $\{\tau\}$ with a closed halfsphere centered at $(\frac{1}{2},0,0)$, see Fig. 1. This follows also from the first assertion of Lemma 3.3, noting that $\Xi^{\mathsf{M},\cap}_E$ is nonempty only for those lines E that intersect C in a halfline and equals the intersection of E with the halfsphere. By Corollary 3.4, $\mathsf{K}^*_G(\tau)$ is the union of $\mathsf{K}_G(\theta)$ over θ in the range $\theta^*_{G,\Xi}(C)$. By Lemma 3.3, $\mathsf{K}_G(\theta) = G$ for $\theta = \tau$, and for θ in the halfsphere, $\mathsf{K}_G(\theta) = \mathsf{N}_\theta(\Xi) \cap [E \setminus G] = E \cap [C \setminus G]$ where E is the line through τ and θ. Hence, $\mathsf{K}^*_G(\tau) = C$ indeed.

5.2. The following result establishes continuity of the GMLE mapping when its domain $\mathsf{dom}(\Psi^*)$ is endowed with the topology that corresponds to the Euclidean topology of the graph of the function Ψ^*, and its range is endowed with the topology of variation distance.

Theorem 5.6. *If a convergent sequence a_n and its limit a are in $\mathsf{dom}(\Psi^*)$, the convergence of $\Psi^*(a_n)$ to $\Psi^*(a)$ implies that $R^*(a_n)$ converges to $R^*(a)$ in variation distance, and also in reversed I-divergence.*

Proof. Let $\epsilon > 0$. Since $\Psi^*(a)$ is finite there exist $\vartheta \in \Xi$ such that

$$\frac{\epsilon^2}{18} > \Psi^*(a) - \big[\langle \vartheta, a\rangle - \Lambda(\vartheta)\big]\,.$$

This and the hypotheses on the convergence imply

$$\frac{2\epsilon^2}{9} > \frac{\epsilon^2}{6} + \Psi^*(a) - \big[\langle \vartheta, a\rangle - \Lambda(\vartheta)\big] > \Psi^*(a_n) - \big[\langle \vartheta, a_n\rangle - \Lambda(\vartheta)\big]\,,$$

the right inequality holding for sufficiently large n. From these inequalities and (2) it follows by the Pinsker inequality $2D(P\|Q) \geqslant |P-Q|^2$, where $|P-Q|$ denotes the variation distance of pm's P,Q, that $\frac{\epsilon}{3} > |R^*(a) - Q_\vartheta|$ and $\frac{2\epsilon}{3} > |R^*(a_n) - Q_\vartheta|$. Thus, the variation distance between $R^*(a)$ and $R^*(a_n)$ is eventually below ϵ, proving that $R^*(a_n)$ converges to $R^*(a)$ in variation distance. The rI-convergence, $D(R^*(a)\|R^*(a_n)) \to 0$, is a consequence of the following lemma. $\square$

Lemma 5.7. *If a variation convergent sequence of pm's P_n and its limit P belong to the range of R^* then $D(P\|P_n) \to 0$.*

Proof. By Theorem 5.1, $P_n = Q_{F_n,\vartheta_n}$ and $P = Q_{F,\vartheta}$ for faces F_n and F of $\mathit{cc}(\mu)$, with $\vartheta_n \in \mathit{cl}(\pi_{F_n}(\Xi))$ and $\vartheta \in \mathit{cl}(\pi_F(\Xi))$. The convergence in variation and [8, Theorem 4] imply that $F_n \supseteq F$ eventually, Q_{F,ϑ_n} converges to P in variation, and $P_n(\mathit{cl}(F)) \to 1$. If $F_n \supseteq F$ holds then $\pi_F(\vartheta_n)$ belongs to $\pi_F(\mathit{cl}(\pi_{F_n}(\Xi))) \subseteq \mathit{cl}(\pi_F(\Xi))$. Therefore, $\pi_F(\vartheta_n) - \vartheta$ is in $\mathit{lin}(\pi_F(\Xi))$ which is a subset of $\mathit{lin}(\Xi) + \mathit{lin}(F)^\perp$. Here, $\mathit{lin}(\Xi) \subseteq M(P)$ by Theorem 5.1, and $\mathit{lin}(F)^\perp \subseteq M(P)$ by Lemma 2.14. Hence, $\pi_F(\vartheta_n) - \vartheta \in M(P)$. It follows that $D(P\|Q_{F,\vartheta_n})$ is eventually finite. By [8, Theorem 3*(i)*], applied to the sequence Q_{F,ϑ_n} converging in variation, $D(P\|Q_{F,\vartheta_n}) \to 0$. This I-divergence equals $D(P\|P_n) + \ln P_n(\mathit{cl}(F))$, and the assertion follows since $P_n(\mathit{cl}(F)) \to 1$. □

Example 5.8. Let μ be sum of the point mass ν at $a = (0,0)$ and the Lebesgue measure on the segment with endpoints $(0,1)$ and $(1,1)$. Then

$$\Lambda(\vartheta) = \ln\left[1 + e^{\vartheta_2} \cdot \tfrac{e^{\vartheta_1}-1}{\vartheta_1}\right], \quad \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2,$$

and Q_ϑ has the mean

$$e^{\vartheta_2} \cdot \left[1 + e^{\vartheta_2} \cdot \tfrac{e^{\vartheta_1}-1}{\vartheta_1}\right]^{-1} \left(\tfrac{\vartheta_1 e^{\vartheta_1} - e^{\vartheta_1} + 1}{\vartheta_1^2}\,,\ \tfrac{e^{\vartheta_1}-1}{\vartheta_1}\right).$$

The pm's Q_{θ_n}, $\theta_n = (-n\,e^n, n)$, in the full exponential family $\mathcal{E}_\mu$, are the GMLE's $R^*(a_n)$ for $a_n = m(Q_{\theta_n})$, and $R^*(a) = \nu$. Then $a_n \to a$ and $D(R^*(a)\|R^*(a_n)) = \Lambda(\theta_n) \to 0$, while $\Lambda^*(\theta_n)$ converges to 1, different from $\Lambda^*(a) = 0$. This shows that the converse of the implication in Theorem 5.6 fails even for a full exponential family with bounded support.

Remark 5.9. If $\mathcal{E} = \mathcal{E}_\mu$ is a full exponential family such that $\mathit{dom}(\Lambda) = \mathbb{R}^d$ then $\mathit{cl}_v(\mathcal{E})$ is equal to the union of the full families $\mathcal{E}_G$ over all faces G of $\mathit{cc}(\mu)$, called the *extension* $\mathit{ext}(\mathcal{E})$ of $\mathcal{E}$ [6]. As all pm's in this extension have a mean, by Theorem 5.1 the range of the GMLE mapping is equal to $\mathit{ext}(\mathcal{E})$ in this case. The barrier cone of $\Xi = \mathit{dom}(\Lambda) = \mathbb{R}^d$ is the singleton $\{0\}$, thus $G^*(a)$ in Theorem 4.1 equals the unique face G of $\mathit{cc}(\mu)$ with $a \in \mathit{ri}(G)$. By Theorems 4.9 and 5.1, $\mathit{dom}(\Lambda^*) = \mathit{cc}(\mu)$ and the mapping $R^* \colon \mathit{cc}(\mu) \to \mathit{ext}(\mathcal{E})$ is bijective. It assigns to each $a \in \mathit{cc}(\mu)$ the unique pm in $\mathit{ext}(\mathcal{E})$ whose mean is equal to a. Note that in the considered case the GMLE can always be identified with a true MLE when $\mathit{ext}(\mathcal{E})$ rather than $\mathcal{E}$ is taken as the model family, see [6, Corollary 12], or the next section. Under the additional assumptions that the set $\mathit{cc}(\mu)$ is bounded and locally simplicial, the stronger result holds that the GMLE mapping is a homeomorphism between $\mathit{cc}(\mu)$ and $\mathit{ext}(\mathcal{E})$, when the latter is endowed with the topology of variation distance. Indeed, boundedness of $\mathit{cc}(\mu)$ implies continuity of the inverse mapping that assigns to $P \in \mathit{ext}(\mathcal{E})$ its mean, the locally simplicial property implies continuity of the function Λ^* on $\mathit{dom}(\Lambda^*) = \mathit{cc}(\mu)$, by [15, Theorem 10.2], and then the mapping $a \mapsto R^*(a)$ is continuous by Theorem 5.6. In the special instance when μ is concentrated on a finite set, the convex hull of this set and $\mathit{ext}(\mathcal{E})$ have been known to be homeomorphic under an MLE with the model family $\mathit{ext}(\mathcal{E})$ [1, Theorem 9.15] and the uniqueness of MLE in $\mathit{ext}(\mathcal{E})$ in terms of means appeared in [1, Theorem 9.16].

5.3. The concept of GMLE is related to that of generalized rI-projection [6]. A set of pm's is *log-convex* [6] if it contains the log-convex combinations of pairs of nonsingular pm's in the set. Here, the log-convex combinations $\overline{P^tQ^{1-t}}$, $0 < t < 1$, of pm's P and Q with μ-densities p and q have the μ-densities $p^tq^{1-t}/\int p^tq^{1-t}d\mu$.

According to [6, Theorem 1], if a set $\mathcal{S}$ of pm's is log-convex and P is any pm such that $D(P\|\mathcal{S}) \triangleq \inf_{Q\in\mathcal{S}} D(P\|Q)$ is finite, there exists a unique pm Q^* that satisfies

$$D(P\|Q) \geqslant D(P\|\mathcal{S}) + D(Q^*\|Q)\,, \qquad Q \in \mathcal{S}\,. \tag{19}$$

This pm Q^* has been called the generalized rI-projection of P to $\mathcal{S}$, because (19) implies variation convergence to Q^* of any sequence of pm's $Q_n \in \mathcal{S}$ with $D(P\|Q_n) \to D(P\|\mathcal{S})$.

The canonically convex exponential families are obviously log-convex, thus the mentioned result applies to $\mathcal{S} = \mathcal{E}_{\Xi}$. In that case, as observed in [6], the generalized rI-projection Q^* of P is equal to the GMLE $R^*(a) = R^*_{\mu,\Xi}(a)$ if the pm P has a mean and the mean equals a. More generally, for any pm P on $\mathbb{R}^d$ with $D(P\|\mathcal{E}_{\Xi})$ finite, Lemma 2.16 implies that the nonempty set Γ of all $\vartheta \in \Xi$ with finite $D(P\|Q_\vartheta)$ is convex, and for $\vartheta \in \Gamma$ the equality

$$D(P\|Q_\vartheta) - D(P\|\mathcal{E}_\Gamma) = \Psi^*_\Gamma(\mathsf{m}(P)) - \big[\, \langle \vartheta, \mathsf{m}(P)\rangle - \Lambda(\vartheta)\,\big]$$

follows from (4) by taking supremum over $\theta \in \Gamma$. In particular, $\Psi^*_\Gamma(\mathsf{m}(P))$ is finite, and (19) for $\mathcal{S} = \mathcal{E}_{\Xi}$ is equivalent to

$$\Psi^*_\Gamma(\mathsf{m}(P)) - \big[\, \langle \vartheta, \mathsf{m}(P)\rangle - \Lambda(\vartheta)\,\big] \geqslant D(Q^*\|Q_\vartheta)\,, \qquad \vartheta \in \Xi_P\,.$$

Comparing this with (2), Q^* equals the GMLE $R^*_{\mu,\Gamma}(\mathsf{m}(P))$ by Remark 1.2. Thus, a generalized rI-projection to a canonically convex exponential family is always equal to a GMLE and can be explicitly described via Theorem 4.1.

The extension $\mathsf{ext}(\mathcal{E})$ of an exponential family $\mathcal{E}$ is always log-convex [6, Theorem 2]. On the other hand, the rI-closure of $\mathcal{E}$, the set of all pm's to which some sequence in $\mathcal{E}$ converges in reversed I-divergence, need not be log-convex [7].

Theorem 5.10. *$\mathsf{cl}_v(\mathcal{E}_{\Xi})$ and the range of the* GMLE *mapping are log-convex.*

Proof. Consider two pm's $Q_{F,\vartheta}$ and $Q_{G,\theta}$ in $\mathsf{cl}_v(\mathcal{E})$ where F, G are Ξ-accessible faces of $\mathsf{cc}(\mu)$, $\vartheta \in \Xi_F$ and $\theta \in \Xi_G$. If these pm's are not mutually singular then the μ-measure of $\mathsf{cl}(F) \cap \mathsf{cl}(G)$ is positive, and then $F \cap G \neq \emptyset$ by Fact 2.10. By Corollary 2.5, $F \cap G$ is a Ξ-accessible face of $\mathsf{cc}(\mu)$.

Since the μ-densities of the two pm's are proportional to $e^{\langle\vartheta,x\rangle}$ on $\mathsf{cl}(F)$, respectively to $e^{\langle\theta,x\rangle}$ on $\mathsf{cl}(G)$, and 0 elsewhere, the μ-density of their log-convex combination is proportional to $e^{\langle t\vartheta+(1-t)\theta,x\rangle}$ on $\mathsf{cl}(F) \cap \mathsf{cl}(G)$ and 0 elsewhere. By Fact 2.10, the log-convex combination is equal to $Q_{F\cap G, t\vartheta+(1-t)\theta}$. This pm does not change if $t\vartheta+(1-t)\theta$ is replaced by its projection to $\mathsf{lin}(F\cap G)$. The projector $\pi_{F\cap G}$ maps ϑ into

$$\pi_{F\cap G}(\mathsf{cl}(\pi_F(\Xi)) \cap \mathsf{dom}(\Lambda_F)) \subseteq \mathsf{cl}(\pi_{F\cap G}(\Xi)) \cap \mathsf{dom}(\Lambda_{F\cap G}) = \Xi_{F\cap G}\,.$$

Analogously, also $\pi_{F\cap G}(\theta)$ is in $\Xi_{F\cap G}$, and thus so is $\pi_{F\cap G}(t\vartheta+(1-t)\theta)$ due to convexity of this set. It follows that the log-convex combinations of $Q_{F,\vartheta}$ and $Q_{G,\theta}$ belong to the subfamily $\mathcal{E}_{F\cap G, \Xi_{F\cap G}}$ of $\mathsf{cl}_v(\mathcal{E}_{\Xi})$.

On account of Theorem 5.1, for the log-convexity of the range it suffices to prove that if $\mathsf{lin}(\Xi)$ is contained in $\mathsf{M}(P)$ and $\mathsf{M}(Q)$ for two nonsingular pm's P and Q then it is contained also in $\mathsf{M}(\overline{P^tQ^{1-t}})$, $0 < t < 1$. This holds because for μ-densities p, q and $f \geqslant 0$

$$t\!\int f\, p\, d\mu + (1-t)\!\int f\, q\, d\mu \geqslant \int f\, p^t q^{1-t}\, d\mu$$

which implies that if f is P- and Q-integrable then it is $\overline{P^tQ^{1-t}}$-integrable as well. $\square$

6. MLE IN THE VARIATION CLOSURE

For $a \in \mathbb{R}^d$ define $\Phi(a) = \Phi_{\mu,\Xi}(a)$ by

$$\Phi(a) = \sup \{ \ell_a(F,\vartheta) \colon F \text{ is a } \Xi\text{-accessible face of } \mathit{cc}(\mu) \text{ and } \vartheta \in \Xi_F \} \tag{20}$$

where $\ell_a(F,\vartheta)$ equals $\langle \vartheta, a\rangle - \Lambda_F(\vartheta)$ if $a \in \mathit{aff}(F)$ and $-\infty$ elsewhere. Note that the pairs (F,ϑ) considered here parameterize bijectively the pm's in the variation closure of $\mathcal{E}_\Xi$. Equivalently,

$$\Phi(a) = \sup \{ \Psi^*_{F,\Xi}(a) \colon F \text{ is a } \Xi\text{-accessible face of } \mathit{cc}(\mu) \text{ with } a \in \mathit{aff}(F) \}\,, \tag{21}$$

because if $a \in \mathit{aff}(F)$, thus $\Xi_{F,a} = \Xi_F$, then $\Psi^*_{F,\Xi}(a) = \Psi^*_{F,\Xi_F}(a)$ by Lemma 2.13 and Corollary 2.12.

If $\Phi(a)$ is finite and the supremum in (20) is actually a maximum, a maximizing pair (F^*,ϑ^*) has a statistical interpretation whenever a equals the mean of an i.i.d. sample from an unknown pm $Q_{F,\vartheta}$ in $\mathit{cl}_v(\mathcal{E}_\Xi)$: (F^*,ϑ^*) is an MLE from the sample of the unknown parameter (F,ϑ). The implicit understanding behind this interpretation is that the μ-density of $Q_{F,\vartheta}$ is equal to $e^{\langle\vartheta,x\rangle-\Lambda_F(\vartheta)}$ on $\mathit{aff}(F)$ and 0 elsewhere. This understanding, though not the same as in previous sections, is legitimate as densities can differ on a set of μ-measure 0, and Fact 2.9 implies by a simple induction that $\mathit{aff}(F) \setminus \mathit{cl}(F)$ has μ-measure 0. An MLE in the present sense can exist only if $a \in \mathit{aff}(\mu)$, for otherwise $\Phi(a) = -\infty$.

Theorem 6.1. *For $a \in \mathit{aff}(\mu)$ the equality $\Phi(a) = \Psi^*(a)$ holds, and the supremum in* (20) *is a maximum if and only if a belongs to the intersection of $\mathit{dom}(\Psi^*)$ and $\mathit{aff}(G)$ for $G = G^*(a)$. In this case, $(G, \theta^*_{G,\Xi}(a))$ is the unique maximizer in* (20).

Interpreting a as the mean of an i.i.d. sample from an unknown pm $Q_{F,\vartheta}$ in $\mathit{cl}_v(\mathcal{E}_\Xi)$ as above, Theorem 6.1 gives a necessary and sufficient condition for the existence of an MLE of the unknown parameter (F,ϑ) from this sample. If the condition holds then the MLE is unique, equals $(G, \theta^*_{G,\Xi}(a))$ with $G = G^*(a)$, and thus parameterizes the GMLE $R^*(a)$, see Theorem 4.1. In case of a full exponential family $\mathcal{E}$ with $\mathit{dom}(\Lambda) = \mathbb{R}^d$, each a in $\mathit{dom}(\Lambda^*) = \mathit{cc}(\mu)$ meets that necessary and sufficient condition and therefore the GMLE for the family $\mathcal{E}$ can be identified with the MLE in $\mathit{cl}_v(\mathcal{E}) = \mathit{ext}(\mathcal{E})$, see Remark 5.9.

The proof of Theorem 6.1 is preceded by three lemmas that complement Theorem 4.1 and give more insight into $\Phi(a)$ expressed through (21).

Lemma 6.2. *If $a \in \mathit{ri}(\mu) + \mathit{bar}(\Xi)$ and F is a proper face of $\mathit{cc}(\mu)$ then $\Psi^*(a) < \Psi^*_{F,\Xi}(a)$.*

Proof. The first assumption implies by Theorem 3.1 that $\Psi^*(a) = \langle\theta, a\rangle - \Lambda(\theta)$ for $\theta = \theta^*(a)$ in $\Xi_{\mu,a} = \mathit{cl}(\pi_{\mu,a}(\Xi)) \cap \mathit{dom}(\Lambda)$. This and the second assumption give that $\Psi^*(a)$ is less than $\langle\theta, a\rangle - \Lambda_F(\theta)$, and thus less than $\Psi^*_{F,\Xi_{\mu,a}}(a)$. The proof is completed by observing that $\Psi^*_{F,\Xi_{\mu,a}}(a)$, $\Psi^*_{F,\pi_{\mu,a}(\Xi)}(a)$ and $\Psi^*_{F,\Xi}(a)$ are equal, by Lemma 2.13 and Corollary 2.12, applied to $\mu^{\mathit{cl}(F)}$ in the role of μ. □

Lemma 6.3. *For $a \in \mathit{dom}(\Psi^*)$ and a face F of $\mathit{cc}(\mu)$, $\Psi^*(a) = \Psi^*_{F,\Xi}(a)$ if and only if F contains $G^*(a)$.*

Proof. If F contains $G = G^*(a)$, then $\Psi^*_{F,\Xi}(a)$ is between $\Psi^*(a)$ and $\Psi^*_{G,\Xi}(a)$ which coincide by Theorem 4.1. If $F \not\supseteq G$ then only the case $a \in F + \mathit{bar}(\Xi)$ is of interest, for otherwise

$\Psi^*_{F,\Xi}(a)$ is infinite by Theorem 4.9 while $\Psi^*(a)$ is finite by the first assumption. Then, by Lemma 2.1, the largest face G' of F with $a \in \mathit{ri}(G') + \mathit{bar}(\Xi)$ is well defined, and it is contained in $G = G^*(a)$ by the definition of the latter. Thus, G' is a subset of $G \cap F$, and therefore a proper face of G. By Lemma 6.2 applied to $\mu^{\mathit{cl}(G)}$ rather than μ, $\Psi^*_{G,\Xi}(a) < \Psi^*_{G',\Xi}(a)$. By Theorem 4.1, $\Psi^*_{G,\Xi}(a) = \Psi^*(a)$ and $\Psi^*_{G',\Xi}(a) = \Psi^*_{F,\Xi}(a)$, thus $\Psi^*(a) < \Psi^*_{F,\Xi}(a)$. □

Lemma 6.4. *If F is a Ξ-accessible face of $\mathit{cc}(\mu)$ and $a \in \mathit{aff}(F)$ then $\Psi^*_{F,\Xi}(a) = \Psi^*(a)$.*

Proof. As always $\Psi^*_{F,\Xi}(a) \geqslant \Psi^*(a)$, it may be assumed that $a \in \mathit{dom}(\Psi^*)$. Then, induction argument is applied on the dimension of $\mathit{aff}(\mu)$. The assertion is trivial in the case $F = \mathit{cc}(\mu)$, in particular, it holds if the dimension is zero.

If the Ξ-accessible face F of $\mathit{cc}(\mu)$ is proper, there exists a nonempty access sequence to F adapted to Ξ. Then, its first element θ belongs to $\mathit{rec}(\pi_\mu(\mathit{ri}(\Xi)))$ and exposes a proper face G of $\mathit{cc}(\mu)$ that contains F as its Ξ-accessible face. It suffices to show that $\Psi^*(a) = \Psi^*_{\mu,\Xi}(a)$ is equal to $\Psi^*_{G,\Xi}(a)$, for then the induction hypothesis applies to $\mu^{\mathit{cl}(G)}$, yielding $\Psi^*_{G,\Xi}(a) = \Psi^*_{F,\Xi}(a)$, and the assertion follows.

To this end, observe that Lemma 4.7 can be applied to a, $\Gamma = \pi_\mu(\Xi)$ and θ. In fact, $a \in \mathit{aff}(F)$ implies $a \in \mathit{aff}(\mu)$, thus $\pi_{\mu,a}(\Xi) = \Gamma$, hence $\Psi^*_{\mu,\Xi}(a) = \Psi^*_{\mu,\Gamma}(a)$ by Corollary 2.12. Thus, $a \in \mathit{dom}(\Psi^*)$ implies $a \in \mathit{dom}(\Psi^*_{\mu,\Gamma})$. Since θ exposes G, and a belongs to $\mathit{aff}(F) \subseteq \mathit{aff}(G)$, the halfspace $H_\leqslant = \{x \in \mathbb{R}^d \colon \langle\theta, x-a\rangle \leqslant 0\}$ contains $\mathit{cs}(\mu)$. The hypothesis $\theta \in \mathit{rec}(\mathit{ri}(\Gamma))$ is satisfied since π_μ commutes with relative interiors. Hence, Lemma 4.7 gives the equality $\Psi^*_{\mu,\Gamma}(a) = \Psi^*_{G,\Gamma}(a)$. Here, the left-hand side equals $\Psi^*_{\mu,\Xi}(a)$ as observed above and, similarly, the right-hand side equals $\Psi^*_{G,\Xi}(a)$, on account of $a \in \mathit{aff}(G)$ and the second assertion of Corollary 2.12 applied to $\mu^{\mathit{cl}(F)}$ in the role of μ. □

Proof of Theorem 6.1. The first assertion immediately follows from Lemma 6.4 and (21). If $a \in \mathit{dom}(\Psi^*)$ then $\Psi^*(a) = \Psi^*_{G,\Xi}(a)$ by Theorem 4.1, where $G = G^*(a)$. If additionally $a \in \mathit{aff}(G)$ then $\Xi_{G,a} = \Xi_G$ thus the maximizer $\theta^*_{G,\Xi}(a)$ of $\langle\theta, a\rangle - \Lambda_G(\theta)$ subject to $\theta \in \Xi_{G,a}$ belongs to Ξ_G. Then, $(G, \theta^*_{G,\Xi}(a))$ is a maximizer in (20).

If (F^*, ϑ^*) attains the maximum in (20) then $a \in \mathit{aff}(F^*)$, $\vartheta^* \in \Xi_{F^*}$ and, using the first assertion, $\langle\vartheta^*, a\rangle - \Lambda_{F^*}(\vartheta^*) = \Psi^*(a)$. Thus, a belongs to $\mathit{dom}(\Psi^*)$. As $\Psi^*(a) = \Psi^*_{F^*,\Xi}(a)$ by Lemma 6.4, and $a \in \mathit{aff}(F^*)$ implies $\Xi_{F^*} = \Xi_{F^*,a}$, it follows by Theorem 3.2 that $\vartheta^* = \theta^*_{F^*,\Xi}(a)$ and $a \in \mathit{ri}(F^*) + \mathit{bar}(\Xi)$. The latter implies by the definition of $G = G^*(a)$ that $F^* \subseteq G$. This and Lemma 6.3 prove that $F^* = G$. Finally, (F^*, ϑ^*) is equal to $(G, \theta^*_{G,\Xi}(a))$, proving also $a \in \mathit{aff}(G)$ and uniqueness of the maximizer. □

7. Appendix

Some concepts of this paper depend on the chosen parametrization of the underlying exponential families, but many are actually invariant. This is discussed in detail here. Let μ and ν be nonzero, σ-finite Borel measures on $\mathbb{R}^d$ whose log-Laplace transforms have nonempty effective domains, and let $\Xi \subseteq \mathit{dom}(\Lambda_\mu)$ and $\Gamma \subseteq \mathit{dom}(\Lambda_\nu)$ be nonempty and convex sets.

Lemma 7.1. *If $Q_{\mu,\xi}$ coincides with $Q_{\nu,\eta}$ for some $\xi \in \mathit{dom}(\Lambda_\mu)$ and $\eta \in \mathit{dom}(\Lambda_\nu)$ then for $\tau = \eta - \xi$ and $t = \Lambda_\nu(\eta) - \Lambda_\mu(\xi)$*

(i) $\Lambda_\mu(\vartheta) + t = \Lambda_\nu(\vartheta + \tau)$, $\vartheta \in \mathbb{R}^d$,

(ii) $\mathit{dom}(\Lambda_\nu) = \mathit{dom}(\Lambda_\mu) + \tau$,
(iii) $Q_{\mu,\vartheta} = Q_{\nu,\vartheta+\tau}$, $\vartheta \in \mathit{dom}(\Lambda_\mu)$,
(iv) $\Lambda^*_\mu(a) = \Lambda^*_\nu(a) - [\,\langle \tau, a\rangle - t\,]$, $a \in \mathbb{R}^d$.

Proof. By assumption,

$$\frac{d\mu}{d\nu}(x) = \frac{dQ_{\nu,\eta}}{d\nu}(x) \Big/ \frac{dQ_{\mu,\xi}}{d\mu}(x) = e^{\langle \eta,x\rangle - \Lambda_\nu(\eta) - [\langle \xi,x\rangle - \Lambda_\mu(\xi)]} = e^{\langle \tau,x\rangle - t}\,,$$

and then, for $\vartheta \in \mathbb{R}^d$

$$\Lambda_\mu(\vartheta) = \ln \int_{\mathbb{R}^d} e^{\langle \vartheta,x\rangle}\,\mu(dx) = \ln \int_{\mathbb{R}^d} e^{\langle \vartheta,x\rangle}\,e^{\langle \tau,x\rangle - t}\,\nu(dx) = \Lambda_\nu(\vartheta+\tau) - t\,,$$

thus *(i)* and *(ii)* hold. For $\vartheta \in \mathit{dom}(\Lambda_\mu)$

$$\ln \frac{dQ_{\mu,\vartheta}}{d\nu}(x) = \ln \frac{dQ_{\mu,\vartheta}}{d\mu}(x) + \ln \frac{d\mu}{d\nu}(x) = \langle \vartheta, x\rangle - \Lambda_\mu(\vartheta) + \langle \tau, x\rangle - t$$

which by *(i)* equals $\langle \vartheta+\tau, x\rangle - \Lambda_\nu(\vartheta+\tau)$. Hence, *(iii)* follows. To see *(iv)*, write

$$\Lambda^*_\mu(a) = \sup_{\vartheta\in\mathbb{R}^d} \big[\,\langle \vartheta, a\rangle - (\Lambda_\nu(\vartheta+\tau) - t)\,\big] = \sup_{\vartheta\in\mathbb{R}^d} \big[\,\langle \vartheta+\tau, a\rangle - \Lambda_\nu(\vartheta+\tau)\,\big] - \langle \tau, a\rangle + t$$

using *(i)*. □

If the families $\mathcal{E}_{\mu,\Xi}$ and $\mathcal{E}_{\nu,\Gamma}$ intersect then $\mathcal{E}_\mu$ and $\mathcal{E}_\nu$ intersect, and then $\mathcal{E}_\mu = \mathcal{E}_\nu$ by *(ii)* and *(iii)*. In addition, $Q_{\mu,\vartheta} = Q_{\nu,\vartheta+\tau}$ holds for all $\vartheta \in \mathit{dom}(\Lambda_\mu)$ with a unique τ in $\mathit{lin}(\mu) = \mathit{lin}(\nu)$. This τ is denoted in the sequel by $\tau^*_{\mu,\nu}$.

Recall that the mapping $\theta^*_{\mu,\mathit{dom}(\Lambda_\mu)}$ is defined on $\mathit{ri}(\mu)$ that equals $\mathit{ri}(\mathit{dom}(\Lambda^*_\mu))$, see (13).

Proposition 7.2. *Suppose $\mathcal{E}_\mu = \mathcal{E}_\nu$.*

(i) For $a \in \mathbb{R}^d$, there exists $\theta \in \mathit{dom}(\Lambda_\mu)$ satisfying $\Lambda^*_\mu(a) = \langle \theta, a\rangle - \Lambda_\mu(\theta)$ if and only if $\vartheta \in \mathit{dom}(\Lambda_\nu)$ exists such that $\Lambda^*_\nu(a) = \langle \vartheta, a\rangle - \Lambda_\nu(\vartheta)$, in which case $Q_{\mu,\theta} = Q_{\nu,\vartheta}$.

(ii) If $a \in \mathit{ri}(\mu) = \mathit{ri}(\nu)$, $\theta = \theta^*_{\mu,\mathit{dom}(\Lambda_\mu)}(a)$ and $\vartheta = \theta^*_{\nu,\mathit{dom}(\Lambda_\nu)}(a)$ then $\vartheta = \theta + \tau^*_{\mu,\nu}$ and $Q_{\mu,\theta} = Q_{\nu,\vartheta}$.

(iii) $\mathit{dom}(\Lambda^*_\mu) = \mathit{dom}(\Lambda^*_\nu)$, and for any a in this set the GMLE*'s $R^*_{\mu,\mathit{dom}(\Lambda_\mu)}(a)$ and $R^*_{\nu,\mathit{dom}(\Lambda_\nu)}(a)$ coincide.*

Proof. The assumption implies that $Q_{\mu,\xi} = Q_{\nu,\eta}$ for some $\xi \in \mathit{dom}(\Lambda_\mu)$ and $\eta \in \mathit{dom}(\Lambda_\nu)$. These parameters are chosen to have $\tau = \eta - \xi$ equal to $\tau^*_{\mu,\nu}$.

(i) By Lemma 7.1 *(i)* and *(iv)*, the equality $\Lambda^*_\mu(a) = \langle \theta, a\rangle - \Lambda_\mu(\theta)$ rewrites to $\Lambda^*_\nu(a) = \langle \theta+\tau, a\rangle - \Lambda_\nu(\theta+\tau)$, which means that θ satisfies the first equality if and only if $\vartheta = \theta+\tau$ satisfies the second one. Here, $\theta \in \mathit{dom}(\Lambda_\mu)$ if and only if $\vartheta \in \mathit{dom}(\Lambda_\nu)$ by Lemma 7.1 *(ii)*. Then *(iii)* of this lemma implies the equality of the pm's.

(ii) As argued in Subsection 3.3, $\Lambda^*_\mu(a) = \langle \theta, a\rangle - \Lambda_\mu(\theta)$ holds for the unique element $\theta = \theta^*_{\mu,\mathit{dom}(\Lambda_\mu)}(a)$ of $\mathit{dom}(\Lambda_\mu) \cap \mathit{lin}(\mu)$. This, $\tau \in \mathit{lin}(\mu)$ and the arguments of the proof of *(i)* imply the assertions.

(iii) By Theorem 4.9, $\mathit{dom}(\Lambda^*_\mu)$ equals $\mathit{cc}(\mu) + \mathit{bar}(\mathit{dom}(\Lambda_\mu))$. The first assertion follows since $\mathit{cc}(\mu) = \mathit{cc}(\nu)$ and the barrier cones coincide on account of Lemma 7.1 *(ii)*. The second one is a special case of Proposition 7.4 *(iii)* with $\Xi = \mathit{dom}(\Lambda_\mu)$ and $\Gamma = \mathit{dom}(\Lambda_\nu)$ since $\mathit{dom}(\Lambda^*_\mu) \subseteq \mathit{aff}(\mu)$. □

Lemma 7.3. *Under the assumption and notations of Lemma 7.1,*

(v) $\tau + \Xi \subseteq \mathsf{dom}(\Lambda_\nu)$,

(vi) $\mathcal{E}_{\mu,\Xi} = \mathcal{E}_{\nu,\tau+\Xi}$,

(vii) $\Psi^*_{\mu,\Xi}(a) = \Psi^*_{\nu,\tau+\Xi}(a) - [\,\langle \tau, a\rangle - t\,]$, $a \in \mathbb{R}^d$,

(viii) $\mathsf{dom}(\Psi^*_{\mu,\Xi}) = \mathsf{dom}(\Psi^*_{\nu,\tau+\Xi})$,

(ix) $R^*_{\mu,\Xi}(a) = R^*_{\nu,\tau+\Xi}(a)$, $a \in \mathsf{dom}(\Psi^*_{\mu,\Xi})$.

Proof. All references here are to Lemma 7.1. Obviously, *(v)* follows from *(ii)*, and then, *(vi)* from *(iii)*. A proof of *(vii)* is a simple variation of that of *(iv)* and is omitted. Clearly, *(viii)* follows from *(vii)*. Finally, combining *(vii)*, *(i)* and *(iii)*, (2) rewrites to

$$\Psi^*_{\nu,\tau+\Xi}(a) + \Lambda_\nu(\vartheta+\tau) - \langle \vartheta+\tau, a\rangle \geqslant D(R^*_{\mu,\Xi}(a)\|Q_{\nu,\vartheta+\tau})\,, \qquad \vartheta \in \Xi\,, \tag{22}$$

where $a \in \mathsf{dom}(\Psi^*_{\mu,\Xi}) = \mathsf{dom}(\Psi^*_{\nu,\tau+\Xi})$ by *(viii)*. The uniqueness of GMLE applied to ν and $\tau + \Xi$ establishes *(ix)*, see Remark 1.2. □

The assertions of Proposition 7.2 extend to general canonically convex exponential families only under the additional assumption that a belongs to $\mathsf{aff}(\mu) = \mathsf{aff}(\nu)$, see also Example 3.5.

Proposition 7.4. *Suppose $\mathcal{E}_{\mu,\Xi} = \mathcal{E}_{\nu,\Gamma}$ and $a \in \mathsf{aff}(\mu)$.*

(i) There exists $\theta \in \Xi$ satisfying $\Psi^*_{\mu,\Xi}(a) = \langle \theta, a\rangle - \Lambda_\mu(\theta)$ if and only if $\vartheta \in \Gamma$ exists such that $\Psi^*_{\nu,\Gamma}(a) = \langle \vartheta, a\rangle - \Lambda_\mu(\vartheta)$, in which case $Q_{\mu,\theta} = Q_{\nu,\vartheta}$.

(ii) The conditions $a \in \mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ and $a \in \mathsf{ri}(\nu) + \mathsf{bar}(\Gamma)$ are equivalent, and imply $\vartheta = \theta + \tau^*_{\mu,\nu}$ and $Q_{\mu,\theta} = Q_{\nu,\vartheta}$ where $\theta = \theta^*_{\mu,\Xi}(a)$ and $\vartheta = \theta^*_{\nu,\Gamma}(a)$.

(iii) The conditions $a \in \mathsf{dom}(\Psi^*_{\Xi})$ and $a \in \mathsf{dom}(\Psi^*_{\nu})$ are equivalent, and imply that $R^*_{\mu,\Xi}(a)$ and $R^*_{\nu,\Gamma}(a)$ coincide.

Proof. By the first assumption, $\mathsf{lin}(\mu) = \mathsf{lin}(\nu)$ and $Q_{\mu,\xi} = Q_{\nu,\eta}$ for some $\xi \in \mathsf{dom}(\Lambda_\mu)$ and $\eta \in \mathsf{dom}(\Lambda_\nu)$ with $\eta - \xi$ equal to $\tau = \tau^*_{\mu,\nu}$.

(i) By Lemma 7.3 *(vi)*, $\mathcal{E}_{\mu,\Xi} = \mathcal{E}_{\nu,\tau+\Xi}$, and thus the first assumption implies that and $\pi_\nu(\tau + \Xi) = \pi_\nu(\Gamma)$. Then, $\Psi^*_{\mu,\Xi}(a) = \langle \theta, a\rangle - \Lambda_\mu(\theta)$ holds with some $\theta \in \Xi$ if and only if $\Psi^*_{\nu,\tau+\Xi}(a) = \langle \theta+\tau, a\rangle - \Lambda_\mu(\theta+\tau)$, on account of Lemma 7.3 *(vii)* and Lemma 7.1 *(i)*. By Corollary 2.12 and $a \in \mathsf{aff}(\nu)$, this is equivalent to $\Psi^*_{\nu,\pi_\nu(\tau+\Xi)}(a) = \langle \theta+\tau, a\rangle - \Lambda_\mu(\theta+\tau)$ and, in turn, to $\Psi^*_{\nu,\Gamma}(a) = \langle \vartheta, a\rangle - \Lambda_\mu(\vartheta)$ with some $\vartheta \in \Gamma$ satisfying $\pi_\nu(\theta+\tau) = \pi_\nu(\vartheta)$. Hence, $\pi_\nu(\theta) = \pi_\nu(\vartheta) - \tau^*_{\mu,\nu}$, and the equality of pm's follows from Lemma 7.1 *(iii)*.

(ii) By $a \in \mathsf{aff}(\mu)$ and Lemma 4.4, a belongs to $\mathsf{ri}(\mu) + \mathsf{bar}(\Xi)$ if and only if it belongs to $\mathsf{ri}(\mu) + \mathsf{bar}(\pi_\mu(\Xi))$. Since $\mathsf{ri}(\mu) = \mathsf{ri}(\nu)$ and the barrier cones of $\pi_\mu(\Xi)$ and $\pi_\nu(\Gamma)$ coincide on account of $\pi_\nu(\tau + \Xi) = \pi_\nu(\Gamma)$, the first assertion obtains by applying again Lemma 4.4 with μ replaced by ν and $a \in \mathsf{aff}(\nu)$. The second assertion is a consequence of $\pi_\nu(\theta) = \pi_\nu(\vartheta) - \tau^*_{\mu,\nu}$, obtained in *(i)* above, and the uniqueness of θ and ϑ in $\mathsf{lin}(\mu)$, by Theorem 3.1.

(iii) The first assertion can be proved analogously to the first assertion of *(ii)* above, by Theorem 4.9 and $\mathsf{cc}(\mu) = \mathsf{cc}(\nu)$. For the equality of GMLE's, rewrite (22) to

$$\Psi^*_{\nu,\Gamma}(a) + \Lambda_\nu(\vartheta) - \langle \vartheta, a\rangle \geqslant D(R^*_{\mu,\Xi}(a)\|Q_{\nu,\vartheta})\,, \qquad \vartheta \in \Gamma\,,$$

using Corollary 2.12, $\pi_\nu(\tau + \Xi) = \pi_\nu(\Gamma)$ and $a \in \mathsf{aff}(\nu)$, and recall the uniqueness of GMLE. □

References

[1] Barndorff-Nielsen, O., *Information and Exponential Families in Statistical Theory.* Wiley, New York, 1978.

[2] Brown, L.D., *Fundamentals of Statistical Exponential Families.* Inst. of Math. Statist. Lecture Notes – Monograph Series, Vol. 9, 1986.

[3] Chentsov, N.N., *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs, Amer. Math. Soc., Providence – Rhode Island, 1982 (Russian original: Nauka, Moscow, 1972).

[4] Csiszár, I. and Matúš, F., Convex cores of measures on $\mathbb{R}^d$, *Studia Sci. Math. Hungar.* **38** (2001) 177–190.

[5] Csiszár, I. and Matúš, F., Information closure of exponential families and generalized maximum likelihood estimates, *Proc. 2002 IEEE Int. Symp. Inform. Theory*, p. 434.

[6] Csiszár, I. and Matúš, F., Information projections revisited, *IEEE Trans. Inform. Theory* **49** (2003) 1474–1490.

[7] Csiszár, I. and Matúš, F., On information closures of exponential families: a counterexample, *IEEE Trans. Inform. Theory* **50** (2004) 922-924.

[8] Csiszár, I. and Matúš, F., Closures of exponential families, *Annals of Probability* **33** (2005) 582–600.

[9] Csiszár, I. and Matúš, F., Generalized maximum likelihood estimates for infinite dimensional exponential families. *Proceedings Prague Stochastics 2006*, Prague, Czech Republic, 288–297.

[10] Eriksson, N., Fienberg, S.E., Rinaldo A., and Sullivant S., Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models. *Journal of Symbolic Computation* **41** (2006) 222-233.

[11] Letac, G., *Lectures on Natural Exponential Families and their Variance Functions.* Monografias de Matemática **50**, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1992.

[12] Lauritzen, S.L., *Graphical Models.* Clarendon Press, Oxford 1996.

[13] Rinaldo, A., On maximum likelihood estimation in log-linear models. Technical Report 833, Department of Statistics, Carnegie Mellon University, 2006.

[14] Rinaldo, A., Computing maximum likelihood estimates in log-linear models. Technical Report 835, Department of Statistics, Carnegie Mellon University, 2006.

[15] Rockafellar, R.T., *Convex Analysis.* Princeton University Press, Princeton 1970.