Statistics and Computing manuscript No.
(will be inserted by the editor)

Computing with Fisher geodesics and extended exponential families

F. Critchley · P. Marriott

Received: date / Accepted: date

Abstract Recent progress using geometry in the design of efficient Markov chain Monte Carlo (MCMC) algorithms have shown the effectiveness of the Fisher Riemannian structure. Furthermore, the theory of the underlying geometry of spaces of statistical models has made an important breakthrough by extending the classical theory on exponential families to their closures, the so-called extended exponential families. This paper looks at the underlying geometry of the Fisher information, in particular its limiting behaviour near boundaries, which illuminates the excellent behaviour of the corresponding geometric MCMC algorithms. Further, the paper shows how Fisher geodesics in extended exponential families smoothly attach the boundaries of extended exponential families to their relative interior. We conjecture that this behaviour could be exploited for trans-dimensional MCMC algorithms.

Keywords Extended exponential family · Fisher geodesic · Fisher metric · Information geometry · Markov chain Monte Carlo

1 Introduction

The seminal paper [11] illustrates a very important way that geometry can have an impact on statistical practice. It considers, under regularity, parameter spaces of statistical models as smooth manifolds, and designs highly efficient Markov Chain Monte Carlo (MCMC) algorithms by imposing (a) Hamiltonian structures and (b) using Langevin methods on Riemmanian geometric structures. The current paper focuses exclusively on the latter. In particular, on the geometry induced – i.e. the geometry induced by the Fisher information metric, [2] – but, critically, looks at wider classes of parametric families by including their closures, thereby including boundaries in the geometry. In many cases it is at, or near, these boundaries, that standard statistical methodologies break down, [3], and this is precisely where geometry can play a very important role.

Consider first a motivating and illustrative example, that of the extended trinomial family.

Example 1 The extended trinomial family is the closure of the trinomial family and so is parameterised by

$$\Delta^2 := \left\{ \pi = (\pi_0, \pi_1, \pi_2)^\top \,:\, \pi_i \geq 0\,, \sum_{i=0}^{2} \pi_i = 1 \right\},$$

where the boundary, in which cell probabilities, π_i, can be exactly zero, is included in the model.

Figure 1 shows this parameter space as a closed simplex in $\mathbb{R}^2$. We consider the Riemannian structure defined by the Fisher information, whose geodesics in this case have a closed form, [12] – although for the figure the geodesic was computed with the same numerical methods used in the other examples. One such (extended) geodesic is plotted (solid red line) and it can be clearly seen that it smoothly touches the boundary. We also note for completeness, that other important (non-metric) geodesics, defined by [2] as the ± 1-geodesics, are also shown. As is generically true the $+1$-geodesic converges to a vertex, while the -1-geodesic cuts the boundary.

Thanks to EPSRC for grants EP/E017878/1 and EP/L010429/1

Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes, Buckinghamshire. MK7 6AA, UK. E-mail: f.critchley@open.ac.uk · Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Ontario, Canada N2L 3G1 E-mail: pmarriot@uwaterloo.ca

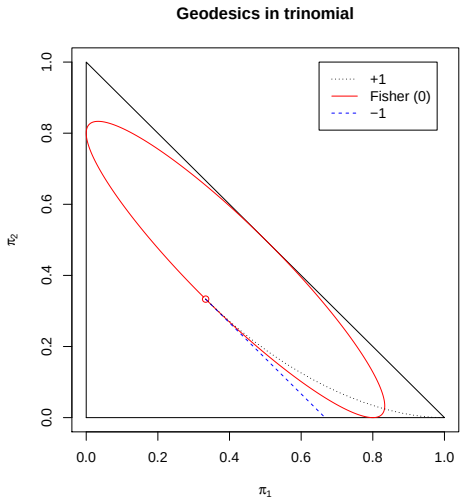

Fig. 1 An extended Fisher geodesic in the extended trinomial family. The extended trinomial family is represented by the closed 2-simplex while a Fisher geodesic is shown by the solid ellipse which smoothly touches all three edges. Two other geodesics, the ± 1, are also shown. The $+1$ converges to a vertex, while the -1 cuts the boundary.

Of key importance, for this paper, are the attractive properties of the Fisher metric – and corresponding geodesics – near boundaries and the excellent performance of MCMC methods. Mostly we take a purely information geometric approach but the final section considers in more detail the links between the Fisher information and the related metrics used in MCMC. The paper is structured as follows. In §2 we look at exponential sub-families of the simplex (Definition 1). This very general class of models includes general finite discrete models, logistic regression, log-linear and other models for categorical data analysis, and exponential family random graph models. These models have also been used as a proxy for a universal class of models, see [7]. Indeed, [4] shows that the information geometry of a continuous model, such as the normal, can be arbitrarily well approximated by a discrete model by discretising the sample space to a fine enough level. Accordingly, there is no real loss in embedding such models inside a sufficiently high dimensional simplex. This approximation is particularly appropriate for Bayesian analyses which condition on finite sets of data. In §3.1 we show how the Fisher information adapts to the boundary of the polytope which defines the mean parameter space and its dual in the natural parameter space. This adaption illuminates the excellent behaviour of the corresponding geometric MCMC algorithms as is shown in §4. Furthermore, we show in §3.2 how the Fisher geodesic smoothly attaches to the boundary, illustrated in Fig. 1, and conjecture that this property may be useful for developing MCMC algorithms in trans-dimensional models, in §5. For brevity, proofs are omitted.

2 Extended multinomial models

Closures of exponential families have been studied by [5], [6], [13] and [8]. The shape of the likelihood function, and hence of the posterior distribution, can be dominated by the behaviour at the boundary. For example, papers [10], [16] and [9] point out that existence, and non-existence, of the maximum likelihood estimate can be fully characterised by considering this closure.

Consider the general k–dimensional extended multinomial model

$$\Delta^k := \left\{ \pi = (\pi_0, \pi_1, \ldots, \pi_k)^T, \pi_i \geq 0\,, \sum_{i=0}^{k} \pi_i = 1 \right\}. \quad (1)$$

The multinomial family on $k+1$ categories can be identified with the (relative) interior of this space, $int(\Delta^k)$, while the extended family, (1), allows the possibility of distributions with different support sets.

Definition 1 The following definitions will be used throughout.

1. The class of models we consider in the paper are all exponential families which are subsets of Δ^k. We call these *exponential sub-families*.
2. A *polytope* in $\mathbb{R}^n$ is the intersection of closed half-spaces. In particular the mean parameter space of an exponential sub-family is the relative interior of a polytope.

Let us consider an example of such an exponential sub-family of Δ^k.

Example 2 A two dimensional extension of the binomial family, as described by Altham in [1], is given by

$$\binom{k}{y} \exp(y\eta + T(y)\phi - \psi(\eta, \phi)) \in \Delta^k, \quad (2)$$

where we take $T(y) = y^2$, $y = 0, \cdots, k$, which is equivalent to Altham's model. This allows both over and under dispersion relative to the binomial and for large k can be thought of as a finite, discrete approximation to the normal family. For a general discussion on the geometry of such discrete approximations see [4], and for more details of the model see the Appendix.

Consider Fig. 2 which represents a mean parameterization of this exponential family for the case $k = 12$. The parameter space is the polytope whose boundary is plotted in blue and it is precisely on the boundary of this polytope where the support of the distribution

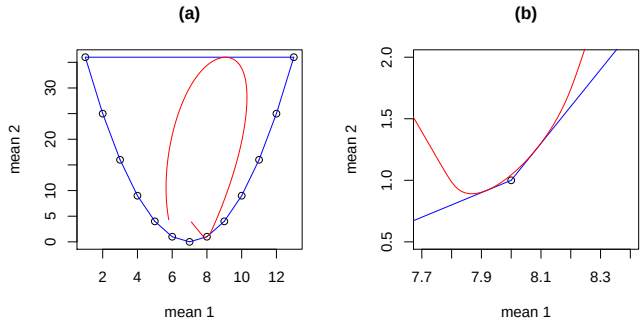

Fig. 2 Fisher geodesic in the multiplicative binomial example. Panel (a) shows the mean parameter space as a convex set in $\mathbb{R}^2$ with the Fisher geodesic (in red) which smoothly touches the boundary. Panel (b) shows a detail of the intersection with the boundary.

changes. The vertices of the polytope are plotted with circles and edges with blue lines. Geometrically this boundary corresponds to the points where the two dimensional model intersects the boundary of Δ^k, see for details [7].

In this plot, in red, we have plotted a Fisher geodesic, computed numerically, shown globally in Panel (a) and with detail of the boundary behaviour in Panel (b) As with Example 1 it can be seen that the geodesic smoothly intersects the boundary of the parameter space.

2.1 The boundary and closures

Before we consider extended exponential families and their boundaries, we recall the definition of the full exponential family when embedded in Δ^k.

Definition 2 Let $\pi^0 = (\pi_i^0) \in int(\Delta^k)$ (i.e. $\pi_i^0 > 0$, $i = 0, \ldots, k$), and V be a $(k+1) \times p$ matrix of the form $(v^{(1)}|\ldots|v^{(p)}) = (v_0|\ldots|v_k)^T$ with linearly independent columns and chosen such that $1_{k+1} := (1, \ldots, 1)^T \notin \text{Range}(V)$. With these definitions there exists a p-dimensional full exponential family in Δ^k, denoted by $\pi(\theta) = \pi_{(\pi^0, V)}(\theta)$ with general element:

$$\begin{aligned}\pi_i(\theta) &= \pi_i^0 \exp\{(V\theta)_i - \psi(\theta)\} \\ &\equiv \pi_i^0 \exp\{v_i^T\theta - \psi(\theta)\}, \end{aligned} \tag{3}$$

$i = 0, \ldots, k$ with normalising constant

$$\exp\{\psi(\theta)\} := \sum_{i=0}^{k} \pi_i^0 \exp\{(V\theta)_i\} = \sum_{i=0}^{k} \pi_i^0 \exp\{v_i^T\theta\},$$

for all $\theta \in \mathbb{R}^p$.

The natural parameter space for Model (3) is simply defined by $\theta \in \mathbb{R}^p$, because for any such θ, $\psi(\theta)$ and hence the corresponding distribution exists. The situation is more complex for the mean parameters defined by

$$\mu = \mu(\theta) = (\mu_1, \ldots, \mu_p)^T = \left(\sum_{i=0}^{k} v_i^{(j)} \pi_i(\theta) \right)_{j=1}^{p}. \tag{4}$$

The range of the mean parameters is the convex hull of the k points $\{v_0, \ldots, v_k\} \subset \mathbb{R}^p$. For example these points, and the corresponding convex hull, are shown in Fig. 2 (a). Thus the parameter space is the relative interior of this polytope.

Definition 3 The extended exponential family corresponds to the closure of Model (3) in the space Δ^k.

Any point in the Model (3) has the property that $\pi_i(\theta) > 0$ for all i. A distribution on the boundary of the extended exponential family, $\bar{\pi}$, correspond to points on the boundary of Δ^k. They are characterised by a change of support, so that there exists at least one $i \in \{0, \ldots, k\}$ such that $\bar{\pi}_i = 0$.

The two parameterisations θ and μ are smooth one to one functions of each other for the full exponential family Model (3), see for example [2]. The manifold based information geometry of Amari, and others [12], is based on the non-linear, but smooth, relationship between these two parameter systems. However as we approach the boundary the degree of non-linearity becomes stronger, and breaks down completely at the boundary; a distribution $\bar{\pi}$ will be uniquely characterised by finite moments but there is no corresponding, finite, natural parameter which corresponding to this distribution. To clarify this difference we look at what happens to the natural parameters at the boundary.

We want to consider the limit points of the p-dimensional full exponential family defined in Def. 2. In particular consider the limiting behaviour for $\lambda \to \infty$ of $\theta = \lambda q$ when $q \in \mathbb{R}^p$, $\lambda \in \mathbb{R}$, and $\|q\| = 1$, i.e. q lies in the sphere in $\mathbb{R}^p$. We can think of each such vector q as a direction in the natural parameter space. From Equation (3) it is clear that the support of the limiting distribution is determined by the maximum elements of the set $\{v_0^T q, \ldots, v_k^T q\}$.

Define $\mathcal{F}_q \subseteq \mathcal{I} := \{0, \ldots, k\}$ such that $i \in \mathcal{F}_q$ if and only if $v_i^T q$ is an upper bound of $\{v_0^T q, \ldots, v_k^T q\}$. In order to characterise limiting distributions consider the convex hull, $conv\{v_0, \ldots, v_k\} \subset \mathbb{R}^p$. For any given unit vector q define the linear function $f_q(v) = v^T q$ over this convex set. The function is maximised on the face defined by vertices $\{v_i | i \in \mathcal{F}_q\}$, the upper bounds of $\{v_0^T q, \ldots, v_k^T q\}$. In other words it is maximised at any point $v = \sum_{i \in \mathcal{F}_q} \rho_i v_i$ where $\sum_{i \in \mathcal{F}_q} \rho_i = 1$ and $\rho_i \geq 0$. This is just the maximum principle for convex functions which states a maximum occurs on the boundary which

is determined by its extremal points corresponding to $\mathcal{F}_q$.

On the face defined by the vertices $\{v_i | i \in \mathcal{F}_q\}$ we have, by definition, that the values $v_i^T q$ are equal for all $i \in \mathcal{F}_q$ hence any point on this face satisfies the linear relationship

$$\sum_{i \in \mathcal{F}_q} \rho_i v_i^T q = b_q$$

where $\sum_{i \in \mathcal{F}_q} \rho_i = 1$ and $\rho_i \geq 0$ and b_q is the attained maximum. Thus we see that q is the normal to a support plane of the convex hull which contains the face.

Thus certain directional vectors in the natural parameter space, q, correspond to limiting distributions whose support is defined by a face of $conv\{v_0, \ldots, v_k\}$. The formal equivalence between the directions in natural parameter space and the convex hull, which lies in the mean parameter space is through the concept of a polar dual of a convex set. This is the set of outward-pointing unit normal vectors to the supporting hyperplanes of the set.

Definition 4 For any set E in R^n the set

$$E^o = \{y \in R^n | \langle y, x \rangle \leq 1, \forall x \in E\}$$

is called the *polar* of E. When E is a closed convex cone, i.e. E is a convex set containing λE for all nonnegative λ, we have

$$E^o = \{y \in R^n | \langle y, x \rangle \leq 0, \forall x \in E\}.$$

For details of the properties of polar sets see Theorem 3 in the Appendix.

Example 3 Returning to Example 2 we see in Fig. 3 Panel (a) the mean parameter space with the polytope of achievable values in blue. The polar of this polytope is shown in Panel (b), also in blue. The vertices of this polar polytope correspond to the directions in the natural parameter space where the limiting distributions correspond to an edge of the polytope in (a). For example if v is a vertex in (b) then

$$\lim_{\theta \to \infty} p(\theta v) = \overline{p}$$

where the moments of $\overline{p}$ lie on the boundary of the polytope in (a). In general $\overline{p}$ will have a smaller support than the distributions in the interior. Such a vector v is called a *direction of recession*.

To illustrate the effect which the boundaries have on the shape of the likelihood function, and hence on any posterior, we have also plotted in Fig. 3 (a) and (b) contours of the log-likelihood (in black). In (a) the log-likelihood is maximised at a point very near the boundary and the shape is very far from quadratic. In

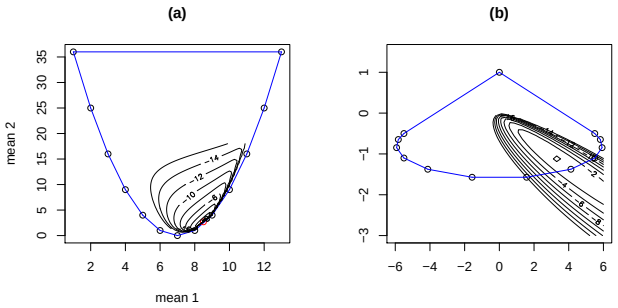

Fig. 3 Likelihood at boundary. Panel (a) shows the mean parameter space for Example 2 and (in black) the contours of the likelihood function. Panel (b) shows the natural parameters for the interior of this model and the polar to the convex set in (a). The contours of the likelihood function are in black.

panel (b) the likelihood is stretched in one of the directions of recession. A small perturbation of the data would put the maximum on the boundary of (a); in (b) the likelihood would then be maximised at a boundary point corresponding to a $\theta \to \infty$ limiting distribution.

3 Fisher geodesics

3.1 Shape of Fisher information

In this section we consider the way that the Fisher geodesic behaves near, and on, the boundary. Formal definitions of geodesics can be found in [2] and [12], and in our Theorem 2 below. First we look at the behaviour of the Fisher information itself.

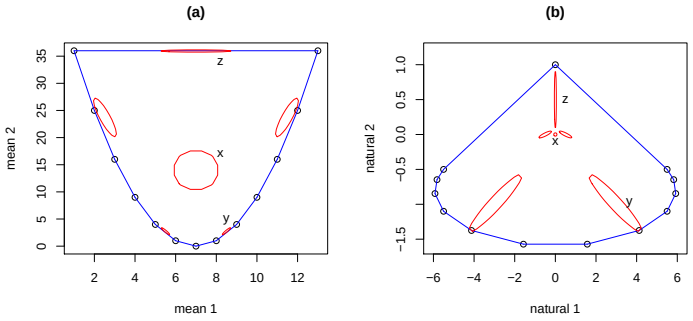

Fig. 4 Fisher geometry in exponential family. Panel (a) uses the mean parameterization with the shape of the Fisher metric, at six points, represented by ellipses. Panel (b) shows the polar and the corresponding Fisher metrics in the natural parameterization.

Example 4 In Fig. 4 we return again to Example 2 and show the boundary polytope, in (a), and its polar, in (b). At a number of points we also plot contours of the functions $\Delta^{(-1)}(\mu,\mu') = (\mu-\mu')^T I(\mu)(\mu-\mu')$ and $\Delta^{(+1)}(\theta,\theta') = (\theta-\theta')^T I(\theta)(\theta-\theta')$ which are based on the Fisher information in the mean and natural parameters respectively. The contours are chosen such that, with a sample size of 100, the distance corresponds to a 99%-percentile from a χ^2_2 random variable, i.e. the ellipses correspond to inferentially relevant regions under a first order asymptotic approximation. Such quadratic measures where used in [3] to determine when the boundary is close enough to effect the adequacy of first order asymptotics.

We see in this example, and in general, that the local geometry, as determined by the Fisher information, changes near the boundary. The contour labelled 'x' is near the center of the model, far from any boundary and is approximately circular in the scale plotted. Contours 'y' and 'z' are much nearer the boundary, in Panel (a), they have adapted their shape such that the dominant eigenvector of the corresponding quadratic form is parallel to the relevant part of the boundary. In contrast, in Panel (b) the corresponding contours are elongated so that the dominant eigen-vector is parallel to the corresponding direction of recession. That in the direction of recession of the relevant vertex of the polar set.

In the example we see that in both parameters the Fisher information, and hence the Riemannian structure, adapts to take into account the boundary, in the mean parameters and the limiting structure at infinity, in the natural. This behaviour can be formalised by the following result.

Theorem 1 *(a) Let μ_i be a sequence of points in the mean parameter space which converge to $\overline{\mu}$, which lies on a face of the boundary polytope, defined by the half space $\langle a,\mu\rangle \leq 1$.*

Let $I(\mu)$ be the Fisher information, $\lambda_{\min}(\mu)$ its minimum eigenvalue and $e_{\min}(\mu)$ the corresponding eigenvector. Then

$$\lim_{i\to\infty} \lambda_{\min}(\mu_i) = 0$$

and $\lim_{i\to\infty} e_{\min}(\mu_i) = a$.

(b) Let θ_i be the corresponding sequence to μ_i in the natural parameters, $I(\theta) := I(\mu(\theta))^{-1}$ the Fisher information, with $\lambda_{\max}(\theta)$ its maximum eigenvalue and $e_{\max}(\mu)$ the corresponding eigenvector. Then

$$\lim_{i\to\infty} \lambda_{\max}(\theta_i) = 0$$

and $\lim_{i\to\infty} e_{\max}(\theta) = a$, which is the vertex in the polar which corresponds to the face in (a).

3.2 Boundary behaviour of Fisher geodesics

Consider again Example 1, where the Fisher geodesic has a simple closed form, [2] or [12]. It can be easily seen, for example in Fig. 1, that the geodesic touches the boundary in a smooth way. In fact since the path shown in the figure intersects the boundary it is more general than a geodesic in a smooth manifold, which would be restricted to the relative interior. This example motivated the following definition.

Definition 5 Let $\mathcal{M}$ be an extended exponential family, which can be written as the union $\bigcup_{i=1}^N \mathcal{M}_i$ of exponential families. A curve $\gamma : [0,T] \to \mathcal{M}$ is called an *extended geodesic* if (a) it is smooth in $\mathcal{M}$ and (b) its restriction to each manifold $\mathcal{M}_i$ is a geodesic.

Example 5 In Example 1 the curve shown Fig. 1 is an extended geodesic since it is the union of geodesics curves which are smoothly connected. It has three components in the relative interior of the simplex – which is an exponential family – and three components in the relative interiors of the three faces – each of which is also an exponential family.

We can characterise geodesics in extended exponential families as being curves of (locally) minimum length in the same way we would in exponential families, with the behaviour of the Fisher information at the boundary (Theorem 1) adding a natural geometric constraint.

It is sufficient to consider paths of (locally) minimum length which connect a point μ^0 in the relative interior of the mean parameter polytope and μ^1 which lies on a face defined by a half space condition $\langle a,\mu\rangle \leq 1$. For a curve connecting these two points to have minimal length, it is of course necessary that this length be finite. This motivates looking at the set of smooth paths

$$\mathcal{P} := \left\{\gamma : [0,1] \to \mathcal{M} | \gamma(0) = \mu^0, \gamma(1) = \mu^1, D(\gamma) < \infty\right\}$$

where

$$D(\gamma) := \int_0^1 \sqrt{\langle \gamma'(s), \gamma'(s)\rangle_{\gamma(s)}} ds < \infty.$$

and defining an extended geodesic to be $\arg\min_{\mathcal{P}} D(\gamma)$.

Theorem 2 *(a) The extended geodesic satisfies the equations*

$$\frac{d^2\mu_i}{dt^2} + \sum_{p,q} \Gamma^i_{pq} \frac{d\mu_p}{dt}\frac{d\mu_q}{dt} = 0 \tag{5}$$

where

$$\Gamma^k_{ij} = \frac{1}{2}\sum_{k,l} I^{kl}\left(\frac{\partial I_{li}}{\partial \mu_j} + \frac{\partial I_{lj}}{\partial \mu_i} - \frac{\partial I_{ij}}{\partial \mu_l}\right),$$

in the relative interior of the natural parameter space. Note that the dependence on θ has been dropped for clarity and $I^{kl} := I_{kl}^{-1}$.

(b) The extended geodesic has the property that $\gamma'(1)$ is tangent to the face containing μ^1.

4 Applications

In this section we look at two of the key examples of [11] where it was shown that the geometric approach to designing MCMC algorithms created excellent results. In Example 6 we look at a (discretised) version of the normal model and in Example 7 we look at logistic regression. Note that [4] shows that the information geometry of a continuous model, such as the normal, can be arbitrarily well approximated by a discrete model by discretising the sample space to a fine enough level. Thus there is no real loss in Example 6 in using the embedding inside a high dimensional simplex.

Example 6 Discretised Normal. In Fig. 5 we see a plot of the mean parameter space for Model 2 where we have chosen a large number of bins relative to the standard deviation of the underlying random variable. Thus this is a finite, discretised version of the normal distribution. The polytope boundary, in the space $\big(E(X), E(X^2)\big)$ is well approximated by the continuous curve defined by

$$Var(X) = E(X^2) - E(X)^2 = 0,$$

as can be clearly seen in the figure. In the plot the red curves are level sets of the Fisher geodesic distance – which has a closed form in the normal model. We see that this distance respects the polytope boundary.

In [11] the normal example was used to demonstrate the way that the Riemannian based MCMC algorithm was very efficient near the boundary. We can see the way the the Fisher metric, and associated Riemannian geometry, adapts to the global geometry of the boundary here.

Example 7 Logistic regression The second example from [11] considered here is the case of logistic regression. The information geometry of this example has been considered in [3] where it was shown how the Fisher metric near the boundary determined the quality of the first order asymptotic approximation. In this example the Fisher geodesic no longer is in closed form, and hence has to be computed numerically. This was done, in the relative interior using the R function *ode* {*deSolve*}, [15], which solves the system of ordinary differential equations given in Theorem 2.

Fig. 6 shows the image of the geodesic, in red, and the boundary of polytope of the mean parameter space.

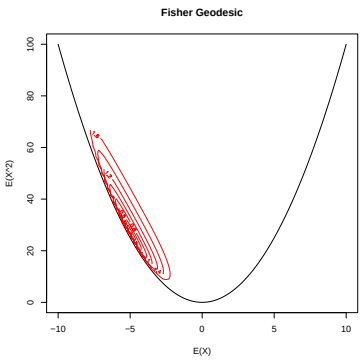

Fig. 5 Geometry of the discretised Normal. The boundary of the mean parameterization for the (discretised) normal is closely approximated by a parabola. The level sets of the Fisher geodesic distance from an interior point are shown in red.

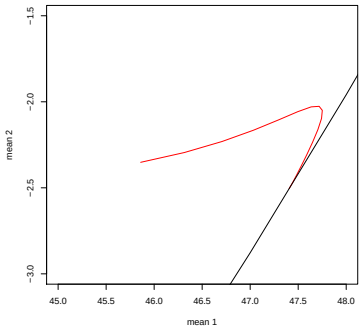

Fig. 6 Geometry of logistic regression: the red line is the Fisher geodesic while the black line is the boundary in the mean parameter space

As can be seen the Fisher geodesic is highly non-linear but does, as predicted, smoothly approach the boundary.

5 Discussion

This paper looks at the Information Geometry – in particular the Fisher Riemannian geometry – of a broad class of extended exponential families, such families meeting their boundaries. Strong non-linearity between the mean and natural parameters in an exponential families corresponds to strong departures from quadratic behaviour for the log-likelihood and corresponding log-posterior, see Fig. 3. Furthermore the strength of non-linearity increases without limit as we approach the boundary of the extended exponential family. This goes some way to explain the success of Riemannian based MCMC methods, since the Fisher metric is the appro-

priate adjustment between the two parameterizations. See in particular Fig. 4.

We note, however, as described by [14], that the metric used in geometric MCMC is an adapted version of the Fisher Information rather than the exact version discussed in this paper. For example [11] proposed using

$$\mathbb{E}_{y|x}\left[-\frac{\partial^2}{\partial x_i \partial x_j}\log f(y|x)\right] - \frac{\partial^2}{\partial x_i \partial x_j}\log \pi_0(x), \qquad (6)$$

where $\pi(x|y) \propto f(y|x)\pi_0(x)$ is the target density for the MCMC, f denotes the likelihood and π_0 the prior. The additional term is designed to allow for the effect of the prior. Note that such a formulation allows the possibly that (6) is not positive definite, and also that it does not transform according to standard metric tensor rules. Nevertheless, it might be hoped that the nice geometry of the Fisher Information might transfer across to geometries defined by (6). We note that it does in commonly used examples, such as the conjugate Dirichlet prior for a multinomial model. There, the adaptation results in a scaled version of the Fisher Information and all properties discussed in this paper would naturally still apply.

For concreteness, consider the simple example of the binomial model where, in the natural parameterisation, θ, the Fisher information is

$$n\pi(\theta)(1-\pi(\theta)) = n\frac{\exp(\theta)}{(1+\exp(\theta))^2}$$

which goes to zero as $\theta \to \pm\infty$. With the (conjugate) Beta(α, β) prior the negative hessian of the log prior is given by

$$(\alpha+\beta-2)\frac{\exp(\theta)}{(1+\exp(\theta))^2}$$

i.e. a scaled version of the Fisher information, with the same boundary behaviour. However, if a normal prior – on the natural parameter scale – were selected, the hessian of its log would not tend to zero as $\theta \to \pm\infty$ and we would see qualitatively different behaviour from (6) at the boundary than that of the Fisher Information. It is therefore an interesting open issue to describe the complete relationship between prior choice, MCMC design and the appropriate geometric behaviour near the boundary.

Note this paper has looked at the behaviour of metrics and their corresponding geodesics near boundaries. An interesting open question, for further investigation, is the extent to which the properties of (extended) geodesics may be directly related to algorithmic performance. The behaviour of the corresponding Fisher geodesic at the boundary, in particular the way it smoothly attaches to the boundary, also points to ways that the Riemannian structure can be further exploited by MCMC methodology. It gives a geometrically, and statistically, natural way to connect models with different support sets. This raises the intriguing possibility that the Fisher geodesic could be used to construct trans-dimensional MCMC algorithms which work across models with different supports.

Appendix

Details of Example 2

The following details of Example 2 may be helpful. In the paper [1] the distribution over $\{0, 1, \ldots, k\}$ is defined, up to proportionality, by

$$P(Z=z; p, \theta) \propto \binom{k}{z} p^z (1-p)^{k-z} \theta^{z(k-z)},$$

where $0 < p < 1$ and $\theta > 0$. This can be written in exponential family form as

$$P(Z=z; p, \theta) \propto \binom{k}{z} \exp\left\{z\log(p/(1-p)) + z(k-z)\log\theta\right\},$$

where the natural parameters are $\log(p/(1-p)), \log\theta$ and sufficient statistics are $z, z(k-z)$. A affine transformation then takes this model to the form used in this paper.

Polar structures

Theorem 3 *(i) The polar E^o of any set E is a closed convex set containing the origin; if $E \subset F$ then $F^o \subset E^o$ (ii) $E \subset E^{oo}$; if E is closed convex and contains the origin the $E^{oo} = E$; (iii) $0 \in intE$ if and only if E^o is bounded.*

The system

$$\langle a^i, x\rangle \leq b_i$$

by shifting the origin to zero and dividing can be written as

$$\langle a^i, x\rangle \leq 1, i = 1, \ldots, p, \qquad (7)$$
$$\langle a^i, x\rangle \leq 0, i = p+1, \ldots, n \qquad (8)$$

Let P be polyhedron define by (7) and (8), then the polar of P is the polyhedron

$$Q = conv\{0; a^1, \ldots, a^p\} + cone\{a^{p+1}, \ldots, a^m\}$$

and conversely, the polar of Q is the polyhedron P.

If P is bounded then $0 \in P^o$ and

$$Q = conv\{a^1, \ldots, a^p\}$$

Proof See [17] Prop 1.21 (page 28).

References

1. Altham, P.: Two generalizations of the binomial distribution. Applied Statist. **27**, 162–167 (1978)
2. Amari, S.I.: Differential-geometrical methods in statistics. Springer-Verlag (1990)
3. Anaya-Izquierdo, K., Critchley, F., Marriott, P.: When are first order asymptotics adequate? a diagnostic. STAT **3**, 17–22 (2014)
4. Anaya-Izquierdo, K., Critchley, F., Marriott, P., Vos, P.: Computational information geometry: foundations. Proc. of Geometric Science of Information 2013, Springer: Lecture Notes in Computer Science pp. 311 – 318 (2013)
5. Barndorff-Nielsen, O.: Information and exponential families in statistical theory. John Wiley & Sons (1978)
6. Brown, L.: Fundamentals of statistical exponential families: with applications in statistical decision theory. Institute of Mathematical Statistics (1986)
7. Critchley, F., Marriott, P.: Computational information geometry in statistics: theory and practice. Entropy, **16** (5), 2454 – 2471, (2014)
8. Csiszar, I., Matus, F.: Closures of exponential families. The Annals of Probability **33**(2), 582–600 (2005)
9. Fienberg, S., Rinaldo, A.: Maximum likelihood estimation in log-linear models: Theory and algorithms. Annals of Statist. **40**, 996–1023 (2012)
10. Geyer, C.J.: Likelihood inference in exponential families and directions of recession. Electron. J. Statist. **3**, 259–289 (2009)
11. Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. J. of the Royal Statistical Society: B **73**(2), 123 – 214 (2011)
12. Kass, R., Vos, P.: Geometrical foundations of asymptotic inference. John Wiley & Sons (1997)
13. Lauritzen, S.: Graphical models. Oxford University Press (1996)
14. Livingstone, S., Girolami, M.: Information-Geometric Markov Chain Monte Carlo Methods Using Diffusions Entropy, **16**, 3074 – 3102 (2014)
15. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2006). URL http://www.R-project.org. ISBN 3-900051-07-0
16. Rinaldo, A., Feinberg, S., Zhou, Y.: On the geometry of discrete exponential families with applications to exponential random graph models. Electron. J. Statist. **3**, 446–484 (2009)
17. Tuy, H.: Convex analysis and global optimization. Klumer academic publishers: London (1998)