Computing the Volume Element of a Family of Metrics on the Multinomial Simplex

Guy Lebanon

June 2, 2003
CMU-CS-03-145

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Keywords: Riemannian geometry, information geometry, metric learning, text classification

Abstract

We compute the differential volume element of a family of metrics on the multinomial simplex. The metric family is composed of pull-backs of the Fisher information metric through a continuous group of transformations. This note complements the paper by Lebanon [3] that describes a metric learning framework and applies the results below to text classification.

1 Basic Concepts from Riemannian Geometry

We start with a brief discussion of some basic concepts from differential geometry and refer to [1] for a more detailed description. A Riemannian metric g, on an nth dimensional differentiable manifold $\mathcal{M}$, is a function that assigns for each point of the manifold $x \in \mathcal{M}$ an inner product on the tangent space $T_x\mathcal{M}$. The metric is required to satisfy the usual inner product properties and to be C^∞ in x.

The metric allows us to measure lengths of tangent vectors $v \in T_x\mathcal{M}$ as $\|v\|_x = \sqrt{g_x(v,v)}$, leading to the definition of a length of a curve on the manifold $c : [a,b] \to \mathcal{M}$ as $\int_a^b \|\dot{c}(t)\| dt$. The geodesic distance function $d(x,y)$ for $x,y \in \mathcal{M}$ is defined as the length of the shortest curve connecting x and y and turns the manifold into a metric space.

For a Riemannian manifold $(\mathcal{M}, g)$ the differential volume element of the metric at $x \in \mathcal{M}$ is given by the square root of the determinant $\text{dvol}g(x) = \sqrt{\det g(x)}$. The volume element $\text{dvol}(x)$ summarizes the size of the metric at x in a scalar. Intuitively, paths crossing areas with high volume will tend to be longer than the same paths over an area with low volume.

Let $F : \mathcal{M} \to \mathcal{N}$ be a diffeomorphism of the manifold $\mathcal{M}$ onto the manifold $\mathcal{N}$. Let $T_x\mathcal{M}, T_y\mathcal{N}$ be the tangent spaces to $\mathcal{M}$ and $\mathcal{N}$ at x and y respectively. Associated with F is the push-forward map F_* that maps $v \in T_x\mathcal{M}$ to $v' \in T_{F(x)}\mathcal{N}$. It is defined as

$$v(h \circ F) = (F_* v)h, \ \forall h \in C^\infty(\mathcal{N}).$$

Intuitively, the push forward maps velocity vectors of curves to velocity vectors of the transformed curves.

Assuming a Riemannian metric h on $\mathcal{N}$, we can obtain a metric F^*h on $\mathcal{M}$ called the pullback metric

$$F^*h_x(u,v) = h_{F(x)}(F_* u, F_* v)$$

where F_* is the push-forward map defined above. The importance of this map is that it turns F (as well as F^{-1}) into an isometry; that is,

$$d_{F^*h}(x,y) = d_h(F(x), F(y)).$$

2 A Family of Metrics on the Simplex

We start by defining the n-simplex by

$$\mathcal{P}_n = \left\{ x \in \mathbb{R}^{n+1} : \forall i, \ x_i \geq 0, \ \sum_{i=1}^{n+1} x_i = 1 \right\}$$

and the n-positive sphere by

$$\mathcal{S}_n^+ = \left\{ x \in \mathbb{R}^{n+1} : \forall i, \ x_i \geq 0, \ \sum_{i=1}^{n+1} x_i^2 = 1 \right\}.$$

The interior of the above manifolds will be dennoted by $\text{int}\mathcal{P}_n$ or $\text{int}\mathcal{S}_n^+$.

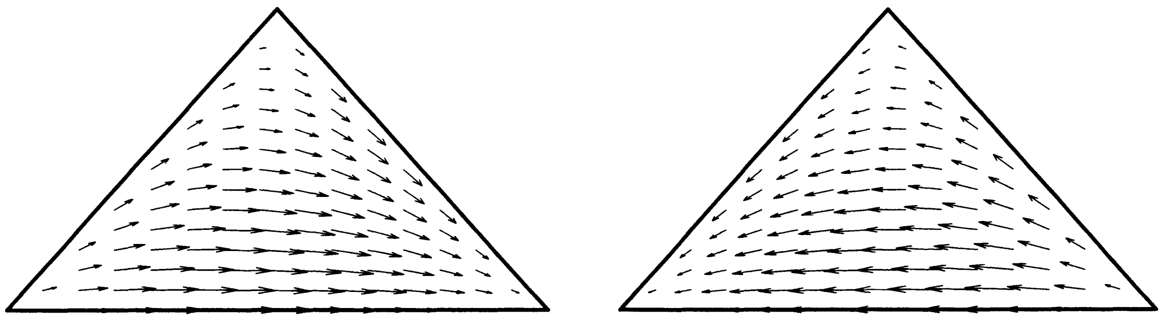

Figure 1: The action of F_λ (left) and F_λ^{-1} (right) on $\mathcal{P}_2$ for $\lambda = (\frac{2}{10}, \frac{5}{10}, \frac{3}{10})$

Consider the following family of diffeomorphisms $F_\lambda : \text{int}\mathcal{P}_n \to \text{int}\mathcal{P}_n$

$$F_\lambda(x) = \left(\frac{x_1\lambda_1}{x\cdot\lambda}, \ldots, \frac{x_{n+1}\lambda_{n+1}}{x\cdot\lambda}\right), \quad \lambda \in \text{int}\mathcal{P}_n$$

where $x \cdot \lambda$ is the scalar product $\sum_{i=1}^{n+1} x_i\lambda_i$. The family F_λ is a Lie group of transformations under composition that is isomorphic to $\text{int}\mathcal{P}_n$. The identity element is $(\frac{1}{n+1}, \ldots, \frac{1}{n+1})$ and the inverse of F_λ is $(F_\lambda)^{-1} = F_\eta$ where $\eta_i = \frac{1/\lambda_i}{\sum_k 1/\lambda_k}$. The above transformation group acts on $x \in \text{int}\mathcal{P}_n$ by increasing the components of x with high λ_i values while remaining in the simplex. See Figure 1 for an illustration of the above action in $\mathcal{P}_2$.

We study the volume properties of metrics on $\mathcal{P}_n$ that are expressed as pull-backs through $F_\lambda^*\mathcal{J}$ of the Fisher information metric $\mathcal{J}$

$$\mathcal{J}_{ij}(x) = \sum_{k=1}^{n+1} \frac{1}{x_k}\frac{\partial x_k}{\partial x_i}\frac{\partial x_k}{\partial x_j}.$$

We now describe a well-known way of characterizing the Fisher information on $\mathcal{P}_n$ as a pull-back metric from the positive n-sphere $\mathcal{S}_n^+$ (see for example [2]). The transformation $R : \mathcal{P}_n \to \mathcal{S}_n^+$ defined by

$$R(x) = (\sqrt{x_1}, \ldots, \sqrt{x_{n+1}})$$

pulls-back the Euclidean metric on the surface of the sphere to the Fisher information on the multinomial simplex. As a result we have that $F_\lambda^*\mathcal{J}$ may also be characterized as the pull back of the metric inherited from the Euclidean space on $\mathcal{S}_n^+$ through

$$\hat{F}_\lambda(x) = \left(\sqrt{\frac{x_1\lambda_1}{x\cdot\lambda}}, \ldots, \sqrt{\frac{x_{n+1}\lambda_{n+1}}{x\cdot\lambda}}\right), \quad \lambda \in \text{int}\mathcal{P}_n.$$

3 The Differential Volume Element of $F_\lambda^* \mathcal{J}$

We start by computing the Gram matrix $[G]_{ij} = F_\lambda^* \mathcal{J}(\partial_i, \partial_j)$ where $\{\partial_i\}_{i=1}^n$ is a basis for $T_x\mathcal{P}_n$ given by the rows of the matrix

$$U = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & 0 & \ddots & 0 & -1 \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \in \mathbb{R}^{n \times n+1}. \tag{1}$$

and computing $\det G$ in Propositions 2-1 below.

Proposition 1. *The matrix* $[G]_{ij} = F_\lambda^* \mathcal{J}(\partial_i, \partial_j)$ *is given by*

$$G = JJ^\top = U(D - \lambda\alpha^\top)(D - \lambda\alpha^\top)^\top U^\top \tag{2}$$

where $D \in \mathbb{R}^{n+1 \times n+1}$ *is a diagonal matrix whose entries are* $[D]_{ii} = \sqrt{\frac{\lambda_i}{x_i}} \frac{1}{2\sqrt{\lambda \cdot x}}$ *and* α *is a column vector given by* $[\alpha]_i = \sqrt{\frac{\lambda_i}{x_i}} \frac{x_i}{2(\lambda \cdot x)^{3/2}}$

Note that all vectors are treated as column vectors and for $\lambda, \alpha \in \mathbb{R}^{n+1}$, $\lambda\alpha^\top \in \mathbb{R}^{n+1 \times n+1}$ is the outer product matrix $[\lambda\alpha^\top]_{ij} = \lambda_i \alpha_j$.

Proof. The jth component of the vector $\hat{F}_{\lambda *} v$ is

$$\begin{aligned} [\hat{F}_{\lambda *} v]_j &= \frac{\mathrm{d}}{\mathrm{d}t} \sqrt{\frac{(x_j + tv_j)\lambda_j}{(x + tv) \cdot \lambda}} \Bigg|_{t=0} \\ &= \frac{1}{2} \frac{v_j \lambda_j}{\sqrt{(x_j + tv_j)\lambda_j} \sqrt{(x + tv) \cdot \lambda}} \Bigg|_{t=0} - \frac{1}{2} \frac{v \cdot \lambda \sqrt{(x_j + tv_j)\lambda_j}}{((x + tv) \cdot \lambda)^{3/2}} \Bigg|_{t=0} \\ &= \frac{1}{2} \frac{v_j \lambda_j}{\sqrt{x_j \lambda_j} \sqrt{x \cdot \lambda}} - \frac{1}{2} \frac{v \cdot \lambda \sqrt{x_j \lambda_j}}{(x \cdot \lambda)^{3/2}}. \end{aligned}$$

Taking the rows of U to be the basis $\{\partial_i\}_{i=1}^n$ for $T_x\mathcal{P}_n$ we have, for $i = 1, \ldots, n$ and $j = 1, \ldots, n+1$,

$$\begin{aligned} [\hat{F}_{\lambda *} \partial_i]_j &= \frac{\lambda_j [\partial_i]_j}{2\sqrt{x_j \lambda_j} \sqrt{x \cdot \lambda}} - \frac{\sqrt{x_j \lambda_j}}{2(x \cdot \lambda)^{3/2}} \partial_i \cdot \lambda \\ &= \frac{\delta_{j,i} - \delta_{j,n+1}}{2\sqrt{x \cdot \lambda}} \sqrt{\frac{\lambda_j}{x_j}} - \frac{\lambda_i - \lambda_{n+1}}{2(x \cdot \lambda)^{3/2}} \sqrt{\frac{\lambda_j}{x_j}} x_j. \end{aligned}$$

If we define $J \in \mathbb{R}^{n \times n+1}$ to be the matrix whose rows are $\{\hat{F}_* \partial_i\}_{i=1}^n$ we have

$$J = U(D - \lambda\alpha^\top).$$

Since the metric $F_\lambda^* \mathcal{J}$ is the pullback of the metric on $\mathcal{S}_n^+$ that is inherited from the Euclidean space through $\hat{F}_\lambda$ we have $[G]_{ij} = \hat{F}_{\lambda *} \partial_i \cdot \hat{F}_{\lambda *} \partial_j$ hence

$$G = JJ^\top = U(D - \lambda\alpha^\top)(D - \lambda\alpha^\top)^\top U^\top.$$

□

Proposition 2. *The determinant of $F_\lambda^* \mathcal{J}$ is*

$$\det F_\lambda^* \mathcal{J} \propto \frac{\prod_{i=1}^{n+1}(\lambda_i/x_i)}{(x\cdot\lambda)^{n+1}}. \tag{3}$$

Proof. We will factor G into a product of square matrices and compute $\det G$ as the product of the determinants of each factor. Note that $G = JJ^\top$ does not qualify as such a factorization since J is not a square matrix.

By factoring a diagonal matrix Λ, $[\Lambda]_{ii} = \sqrt{\frac{\lambda_i}{x_i}}\frac{1}{2\sqrt{x\cdot\lambda}}$ from $D - \lambda\alpha^\top$ we have

$$J = U\left(I - \frac{\lambda x^\top}{x\cdot\lambda}\right)\Lambda \tag{4}$$

$$G = U\left(I - \frac{\lambda x^\top}{x\cdot\lambda}\right)\Lambda^2\left(I - \frac{\lambda x^\top}{x\cdot\lambda}\right)^\top U^\top. \tag{5}$$

We proceed by studying the eigenvalues and eigenvectors of $I - \frac{\lambda x^\top}{x\cdot\lambda}$ in order to simplify (5) via an eigenvalue decomposition. First note that if (v,μ) is an eigenvector-eigenvalue pair of $\frac{\lambda x^\top}{x\cdot\lambda}$ then $(v, 1-\mu)$ is an eigenvector-eigenvalue pair of $I - \frac{\lambda x^\top}{x\cdot\lambda}$. Next, note that vectors v such that $x^\top v = 0$ are eigenvectors of $\frac{\lambda x^\top}{x\cdot\lambda}$ with eigenvalue 0. Hence they are also eigenvectors of $I - \frac{\lambda x^\top}{x\cdot\lambda}$ with eigenvalue 1. There are n such independent vectors $v_1,\ldots,v_n$. Since $\text{trace}(I - \frac{\lambda x^\top}{x\cdot\lambda}) = n$, the sum of the eigenvalues is also n and we may conclude that the last of the $n+1$ eigenvalues is 0.

The eigenvectors of $I - \frac{\lambda x^\top}{x\cdot\lambda}$ may be written in several ways. One possibility is as the columns of the following matrix

$$V = \begin{pmatrix} -\frac{x_2}{x_1} & -\frac{x_3}{x_1} & \cdots & -\frac{x_{n+1}}{x_1} & \lambda_1 \\ 1 & 0 & \cdots & 0 & \lambda_2 \\ 0 & 1 & \cdots & 0 & \lambda_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_{n+1} \end{pmatrix} \in \mathbb{R}^{n+1\times n+1}$$

where the first n columns are the eigenvectors that correspond to unit eigenvalues and the last eigenvector corresponds to a 0 eigenvalue.

Using the above eigenvector decomposition we have $I - \frac{\lambda x^\top}{x\cdot\lambda} = V\tilde{I}V^{-1}$ and $\tilde{I}$ is a diagonal matrix containing all the eigenvalues. Since the diagonal of $\tilde{I}$ is $(1,1,\ldots,1,0)$ we may write $I - \frac{\lambda x^\top}{x\cdot\lambda} = V^{|n}V^{-1|n}$ where $V^{|n} \in \mathbb{R}^{n+1\times n}$ is V with the last column removed and $V^{-1|n} \in \mathbb{R}^{n\times n+1}$ is V^{-1} with the last row removed.

We have then,

$$\begin{aligned} \det G &= \det(U(V^{|n}V^{-1|n})\Lambda^2(V^{-1|n\top}V^{|n\top})U^\top) \\ &= \det((UV^{|n})(V^{-1|n}\Lambda^2V^{-1|n\top})(V^{|n\top}U^\top)) \\ &= (\det(UV^{|n}))^2\ \det(V^{-1|n}\Lambda^2V^{-1|n\top}). \end{aligned}$$

Noting that

$$UV^{|n} = \begin{pmatrix} -\frac{x_2}{x_1} & -\frac{x_3}{x_1} & \cdots & -\frac{x_n}{x_1} & -\frac{x_{n+1}}{x_1} - 1 \\ 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \in \mathbb{R}^{n \times n}$$

we factor $1/x_1$ from the first row and add columns $2, \dots, n$ to column 1 thus obtaining

$$\begin{pmatrix} -\sum_{i=1}^{n+1} x_i & -x_3 & \cdots & -x_n & -x_{n+1} - x_1 1 \\ 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

Computing the determinant by minor expansion of the first column we obtain

$$\det(UV^{|n})^2 = \left(\frac{1}{x_1} \sum_{i=1}^{n+1} x_i \right)^2 = \frac{1}{x_1^2}. \tag{6}$$

We now turn to computing $\det V^{-1|n} \Lambda^2 V^{-1|n\top}$. The inverse of V, as may be easily verified is,

$$V^{-1} = \frac{1}{x \cdot \lambda} \begin{pmatrix} -x_1\lambda_2 & x \cdot \lambda - x_2\lambda_2 & -x_3\lambda_2 & \cdots & -x_{n+1}\lambda_2 \\ -x_1\lambda_3 & -x_2\lambda_3 & x \cdot \lambda - x_3\lambda_3 & \cdots & -x_{n+1}\lambda_3 \\ \vdots & \vdots & & \ddots & \\ -x_1\lambda_{n+1} & -x_2\lambda_{n+1} & \cdots & \cdots & x \cdot \lambda - x_{n+1}\lambda_{n+1} \\ x_1\lambda_1 & x_2\lambda_1 & \cdots & \cdots & x_{n+1}\lambda_1 \end{pmatrix}.$$

Removing the last row gives

$$\begin{aligned} V^{-1|n} &= \frac{1}{x \cdot \lambda} \begin{pmatrix} -x_1\lambda_2 & x \cdot \lambda - x_2\lambda_2 & -x_3\lambda_2 & \cdots & -x_{n+1}\lambda_2 \\ -x_1\lambda_3 & -x_2\lambda_3 & x \cdot \lambda - x_3\lambda_3 & \cdots & -x_{n+1}\lambda_3 \\ \vdots & \vdots & & \ddots & \\ -x_1\lambda_{n+1} & -x_2\lambda_{n+1} & \cdots & \cdots & x \cdot \lambda - x_{n+1}\lambda_{n+1} \end{pmatrix} \\ &= \frac{1}{x \cdot \lambda} P \begin{pmatrix} -x_1 & x \cdot \lambda/\lambda_2 - x_2 & -x_3 & \cdots & -x_{n+1} \\ -x_1 & -x_2 & x \cdot \lambda/\lambda_3 - x_3 & \cdots & -x_{n+1} \\ \vdots & \vdots & & \ddots & \\ -x_1 & -x_2 & \cdots & \cdots & x \cdot \lambda/\lambda_{n+1} - x_{n+1} \end{pmatrix}. \end{aligned}$$

where

$$P = \begin{pmatrix} \lambda_2 & 0 & \cdots & 0 \\ 0 & \lambda_3 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_{n+1} \end{pmatrix}.$$

$[V_n^{-1}\Lambda^2 V_n^{-1\top}]_{ij}$ is the scalar product of the ith and jth rows of the following matrix

$$V_n^{-1}\Lambda = \frac{1}{2}(x\cdot\lambda)^{-3/2}P$$

$$\begin{pmatrix} -\sqrt{x_1\lambda_1} & x\cdot\lambda/\sqrt{x_2\lambda_2}-\sqrt{x_2\lambda_2} & -\sqrt{x_3\lambda_3} & \cdots & -\sqrt{x_{n+1}\lambda_{n+1}} \\ -\sqrt{x_1\lambda_1} & -\sqrt{x_2\lambda_2} & x\cdot\lambda/\sqrt{x_3\lambda_3}-\sqrt{x_3\lambda_3} & \cdots & -\sqrt{x_{n+1}\lambda_{n+1}} \\ \vdots & \vdots & & \ddots & \\ -\sqrt{x_1\lambda_1} & -\sqrt{x_2\lambda_2} & \cdots & \cdots & x\cdot\lambda/\sqrt{x_{n+1}\lambda_{n+1}}-\sqrt{x_{n+1}\lambda_{n+1}} \end{pmatrix}.$$

We therefore have

$$V_n^{-1}\Lambda^2 V_n^{-1\top} = \frac{1}{4}(x\cdot\lambda)^{-2}PQP$$

where

$$Q = \begin{pmatrix} \frac{x\cdot\lambda}{x_2\lambda_2}-1 & -1 & \cdots & -1 \\ -1 & \frac{x\cdot\lambda}{x_3\lambda_3}-1 & \cdots & -1 \\ \vdots & & \ddots & \vdots \\ -1 & -1 & -1 & \frac{x\cdot\lambda}{x_{n+1}\lambda_{n+1}}-1 \end{pmatrix}.$$

As a consequence of Lemma 2 in the appendix we have

$$\det Q = x_1\lambda_1 \frac{(x\cdot\lambda)^n}{\prod_{i=1}^{n+1} x_i\lambda_i} - x_1\lambda_1 \frac{(x\cdot\lambda)^{n-1}\sum_{j=2}^{n+1} x_j\lambda_j}{\prod_{i=1}^{n+1} x_i\lambda_i} = x_1^2\lambda_1^2 \frac{(x\cdot\lambda)^{n-1}}{\prod_{i=1}^{n+1} x_i\lambda_i}.$$

The determinant then is

$$\det V_n^{-1}\Lambda^2 V_n^{-1\top} = (1/4)^n (x\cdot\lambda)^{-2n}\left(\prod_{i=2}^{n+1}\lambda_i\right) x_1^2\lambda_1^2 \frac{(x\cdot\lambda)^{n-1}}{\prod_{i=1}^{n+1} x_i\lambda_i}\left(\prod_{i=2}^{n+1}\lambda_i\right) = \frac{x_1^2(x\cdot\lambda)^{n-1}}{4^n(x\cdot\lambda)^{2n}}\prod_{i=1}^{n+1}\frac{\lambda_i}{x_i}$$

The determinant of G is

$$\det G = (\det UV_n)^2 \det V_n^{-1}\Lambda^2 V_n^{-1\top} = \frac{1}{x_1^2}\frac{x_1^2(x\cdot\lambda)^{n-1}}{4^n(x\cdot\lambda)^{2n}}\prod_{i=1}^{n+1}\frac{\lambda_i}{x_i} \propto \frac{\prod_{i=1}^{n+1}(\lambda_i/x_i)}{(x\cdot\lambda)^{n+1}}.$$

□

Note that the determinant does not depend on the choice of the basis for $T_x\mathcal{P}_n$ and is symmetric in all $n+1$ variables

Acknowledgements

I thank John Lafferty and Leonid Kontorovich for interesting discussions and helpful comments.

Appendix

A The Determinant of a Diagonal Matrix plus a Constant Matrix

We prove some basic results concerning the determainants of a diagonal matrix plus a constant matrix. These results are useful in proving Proposition 1.

The determinant of a matrix $\det A \in \mathbb{R}^{n\times n}$ may be seen as a function of the rows of A, $\{A_i\}_{i=1}^n$

$$f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R} \qquad f(A_1, \ldots, A_n) = \det A.$$

The multilinearity property of the determinant means that the function f above is linear in each of its components

$$\forall j = 1, \ldots, n \quad f(A_1, \ldots, A_{j-1}, A_j + B_j, A_{j+1}, \ldots, A_n) = f(A_1, \ldots, A_{j-1}, A_j, A_{j+1}, \ldots, A_n) + f(A_1, \ldots, A_{j-1}, B_j, A_{j+1}, \ldots, A_n).$$

Lemma 1. *Let $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix with $D_{11} = 0$ and $\mathbf{1}$ a matrix of ones. Then*

$$\det(D - \mathbf{1}) = -\prod_{i=2}^{m} D_{ii}.$$

Proof. Subtract the first row from all the other rows to obtain

$$\begin{pmatrix} -1 & -1 & \cdots & -1 \\ 0 & D_{22} & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & D_{mm} \end{pmatrix}.$$

Now compute the determinant by the cofactor expansion along the first column to obtain

$$\det(D - \mathbf{1}) = (-1)\prod_{j=2}^{m} D_{jj} + 0 + 0 + \cdots + 0.$$

□

Lemma 2. *Let $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix and $\mathbf{1}$ a matrix of ones. Then*

$$\det(D - \mathbf{1}) = \prod_{i=1}^{m} D_{ii} - \sum_{i=1}^{m} \prod_{j\neq i} D_{jj}.$$

Proof. Using the multilinearity property of the determinant we separate the first row of $D - \mathbf{1}$ as $(D_{11}, 0, \ldots, 0) + (-1, \ldots, -1)$. The determinant $\det D - \mathbf{1}$ then becomes $\det A + \det B$ where A is $D - \mathbf{1}$ with the first row replaced by $(D_{11}, 0, \ldots, 0)$ and B is the $D - \mathbf{1}$ with the first row replaced by a vector or -1.

Using Lemma 1 we have $\det B = -\prod_{j=2}^{n} D_{jj}$. The determinant $\det A$ may be expanded along the first row resulting in $\det A = D_{11}M_{11}$ where M_{11} is the minor resulting from deleting the first row and the first column. Note that M_{11} is the determinant of a matrix similar to $D - \mathbf{1}$ but of size $n - 1 \times n - 1$.

Repeating recursively the above multilinearity argument we have

$$\begin{aligned}\det(D-\mathbf{1}) &= -\prod_{j=2}^{n} D_{jj} + D_{11}\left(-\prod_{j=3}^{n} D_{jj} + D_{22}\left(-\prod_{j=4}^{n} D_{jj} + D_{33}\left(-\prod_{j=5}^{n} D_{jj} + D_{44}(\cdots)\right)\right)\right) \\ &= \prod_{i=1}^{n} D_{ii} - \sum_{i=1}^{n}\prod_{j\neq i} D_{jj}.\end{aligned}$$

□

References

[1] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 2003.

[2] R. E. Kass. The geometry of asymptotic inference. *Statistical Science*, 4(3):188–234, 1989.

[3] G. Lebanon. Learning riemannian metrics. In *Proc. of the 19th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann publishers, 2003.