The Jacobian of the Fréchet mean on SPD spaces: Implementation details

Mathis Birken

March 17, 2023

We use the same notation as in the other document.

The gradient

Let $X \in \mathrm{SPD}_d$ have orthogonal diagonalization $X = V\Lambda V^T$, where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_d)$. For the derivative of d_I^2 we notice the fact that we can write:

$$\mathrm{d}_X(d_I^2)(Z) = \sum_{i=1}^{d} 2\frac{\log \lambda_i}{\lambda_i}\langle Zv_i, v_i\rangle = 2\,\mathrm{tr}(V^T ZV \underbrace{\mathrm{diag}(..., \frac{\log \lambda_i}{\lambda_i}, ...)}_{=:\tilde{\Lambda}}),$$

because $\langle Zv_i, v_j\rangle = (V^T ZV)_{ij}$.
Now for d_Y^2 we use the chain rule. But because in general $Y^{-1}X$ won't be symmetric, writing $d_Y^2(X) = d_I^2(Y^{-1}X)$ doesn't make sense. Instead, we use conjugation: $d_Y^2(X) = d_I^2(\sqrt{Y^{-1}}X\sqrt{Y^{-1}})$. We realize that conjugation is linear in X and so the chain rule gives the very simple formula

$$\mathrm{d}_X(d_Y^2)(Z) = \mathrm{d}_{\sqrt{Y^{-1}}X\sqrt{Y^{-1}}}(d_I^2)(\sqrt{Y^{-1}}Z\sqrt{Y^{-1}})$$

.

The Hessian

The formula given in the other document contains two small errors. After correcting them, we can write:

$$\begin{aligned}
\mathrm{Hess}_X^D(d_I^2)(Z, W) &= 2\sum_{i=1}^{d} \frac{1 - \log(\lambda_i)}{\lambda_i^2}\langle Zv_i, v_i\rangle\langle Wv_i, v_i\rangle \\
&\quad + 2\sum_{j=i+1}^{d} \left(\frac{\log \lambda_i}{\lambda_i} - \frac{\log \lambda_j}{\lambda_j}\right)\frac{1}{\lambda_i - \lambda_j}\langle Zv_i, v_j\rangle\langle Wv_i, v_j\rangle \\
&= \sum_{i,j=1}^{d} h(\lambda_i, \lambda_j)\langle Zv_i, v_j\rangle\langle Wv_i, v_j\rangle \\
&= \sum_{i,j=1}^{d} H_{ij}(V^T ZV)_{ij}(V^T WV)_{ij} \qquad (H_{ij} := h(\lambda_i, \lambda_j))
\end{aligned}$$

where we define $h : \mathbb{R}^2_{>0} \to \mathbb{R}$ to be the continous function

$$h(x,y) := \begin{cases} \left(\frac{\log x}{x} - \frac{\log y}{y}\right)\frac{1}{x-y}, & x \neq y \\ \frac{1-\log x}{x^2}, & x = y. \end{cases}$$

Now the above sum can be interpreted as the sum over the Hadamard product of three matrices. Equivalently, if we represent those matrices by vectors in $\mathbb{R}^{d^2}$, it is simply the inner product of the Hadamard product $H \circ V^T Z V$ and $V^T W V$.

Let's now turn to the more general case d_Y^2. By a similar argument as above involving the linearity of the conjugation in one argument we have:

$$\mathrm{Hess}_X(d_Y^2)(Z, W) = \mathrm{Hess}_{c(X)}(d_I^2)(c(Z), c(W)) \quad \text{where } c : A \mapsto \sqrt{Y^{-1}}A\sqrt{Y^{-1}}$$

In practice, we want to calculate the Hessian matrix in coordinates, so we plug in our matrix basis vectors $(\partial_i)_{1 \leq i \leq d(d+1)/2}$ for Z and W. If we define the $(d(d+1)/2 \times d^2)$-matrix M to have as rows the reshaped matrices $V^T c(\partial_i) V$, we can perform the calculation above for all ∂_i, ∂_j at once:

$$\mathrm{Hess}_X^D(d_Y^2)(\partial_i, \partial_j) = ((H \circ M)M^T)_{ij}.$$

(The Hadamard product is meant to be applied in each row of M.)

The mixed second derivatives

Now we want to implement the $B_{x,y}$ matrices from the other document. In the notation of the paragraph in practice, in order to calculate the second partial derivatives we need to find a matrix Z such that $\bar{Z}_Y = \partial_j$, for each j. We simply set $Z := \frac{1}{2}\partial_j Y^{-1}$. Now the reasoning given appeerently also works using the flat connection D. (I don't understand the part about the Killing vector field very well unfortunately.) Using $(D_{\partial_i}\bar{Z})_X = Z\partial_i + \partial_i Z^T$ we obtain:

$$\begin{aligned} B_{X,Y}(\partial_i, \partial_j) = &- \mathrm{d}_X(d_Y^2)\left(\frac{1}{2}\partial_i Y^{-1}\partial_j + \frac{1}{2}(\partial_i Y^{-1}\partial_j)^T\right) \\ &- \mathrm{Hess}_X^D(d_Y^2)\left(\partial_i, \frac{1}{2}\partial_j Y^{-1}X + \frac{1}{2}(\partial_j Y^{-1}X)^T\right) \end{aligned}$$

Now we compute these two terms efficiently for all i, j at once. For the first term, we can make use of the equation in terms of the trace. (We use that cyclic permutation under the trace are permissible.)

$$\begin{aligned} \mathrm{d}_X(d_Y^2)(\partial_i Y^{-1}\partial_j) &= 2\,\mathrm{tr}(V^T\sqrt{Y^{-1}}\partial_i Y^{-1}\partial_j\sqrt{Y^{-1}}V\tilde{\Lambda}) \\ &= 2\,\mathrm{tr}(\partial_i Y^{-1}\partial_j\sqrt{Y^{-1}}V\tilde{\Lambda}V^T\sqrt{Y^{-1}}) \end{aligned}$$

Now the key to vectorizing this equation is the fact that the a trace is the same as the inner product of the matrices, reshaped into a vector (with one transpose). Therefore we can use a procedure very similar to the one described in the part about the Hessian and realize the calculation of those inner products in one matrix multiplication. It is a very easy to see that symmetrizing the obtained matrix gets us the desired first term (if we think about the above as a matrix equation.)

For the second term the procedure is quite easy, we can simply turn all the $\frac{1}{2}\partial_j Y^{-1} X + \frac{1}{2}(\partial_j Y^{-1} X)^T$ into row coordinate vectors and stack them to a matrix, that we subsequently multiply by the Hessian maxtrix. This is due to the simple ∂_i in the first argument.