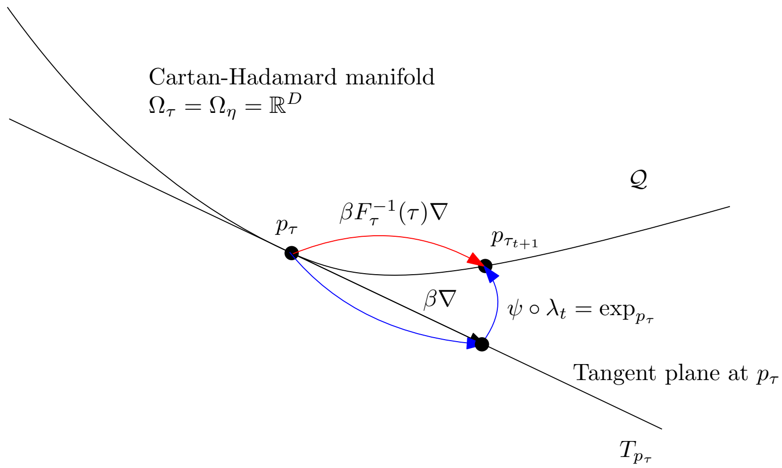

Figure 1: Geometric interpretation of local natural coordinates on a Cartan-Hadamard manifold.

Local parameterizations

Let $L_\tau(\tau) := E_{q(w|\tau)}[\ell(w)]$. Ordinary NGD with respect to global parameter τ is

$$\tau_{t+1} = \tau_t - \beta F_\tau^{-1}(\tau_t)\nabla_\tau L_\tau(\tau_t),$$

where β is the learning rate. Using local parameterization η and auxiliary parameterization λ_t with $\tau_\eta = \psi \circ \phi_{\lambda_t}$ so that $\tau_\eta(\eta) = \tau$ at iteration t via λ_t, we have one iteration of NGD with respect to local parameterization η as:

$$\eta' = \eta_0 - \beta F_\eta^{-1}(\eta_0)\nabla_\eta L_\eta(\eta_0)$$

Since we choose $\eta_0 = 0$, and we have $L_\eta = L_\tau \circ \tau_\eta$, we get using the chain rule (a multivariate generalization of $f(g(y))' = g'(y)f'(g(y))$ with $x = g(y)$):

$$\begin{aligned}\eta' &= -\beta F_\eta^{-1}(\eta_0)\left(\nabla_\eta(L_\tau \circ \tau_\eta)(\eta_0)\right)\\ &= -\beta F_\eta^{-1}(\eta_0)\left(\nabla_\eta(\psi \circ \phi_{\lambda_t})(\eta_0)\right)\nabla_\tau L_\tau(\tau)\end{aligned}$$

Then we map back η' to τ as

$$\tau_{t+1} = \tau_\eta(\eta') = \psi \circ \phi_{\lambda_t}\left(-\beta F_\eta^{-1}(\eta_0)(\nabla_\eta(\psi \circ \phi_{\lambda_t})(\eta_0))\nabla_\tau L_\tau(\tau)\right)$$

0.1 The case of Cartan-Hadamard manifolds

When the Fisher-Rao D-dimensional manifold $\mathcal{Q}$ is a Cartan-Hadamard manifold (non-positive sectional curvature like the manifold of positive-definite matrices) then by Cartan?Hadamard theorem the manifold is diffeomorphic to $\mathbb{R}^D$. In a Cartan-Hadamard manifold, the Riemannian exponential map $\exp_p : T_p\mathcal{Q} \to \mathcal{Q}$ is a covering map. Thus we can choose η to be the Euclidean coordinates in the tangent plane T_p with ordinary gradient, and define $\psi \circ \lambda_t(\eta) = \exp_{p_{\tau_t}}(v)$ to be the Riemannian exponential map. Therefore on any Fisher-Rao manifold which is of Cartan type (e.g., manifold of Gaussians), we can introduce local natural coordinates. Notice that when the FIM is constant, $F_\eta^{-1}(\eta_0)$, then by a change of variable using Cholesky decomposition, we can make it Euclidean. Figure 1 gives a geometric intepretation of the local natural coordinates on such manifolds.