Appendix 1

FACTORIZATION OF PHYLOGENETIC LIKELIHOOD UNDER MULTIVARIATE BROWNIAN MOTION

Below can be found a short derivation of one decomposition for the phylogenetic multivariate Brownian likelihood function, used internally to accelerate the calculation of phylogenetic likelihoods in RevBayes and other software. It involves two separate factorizations that allow for a very high dimensional multivariate normal density to be expressed as the product of univariate normal densities. These factorizations exploit properties of the multivariate Brownian motion rate matrix R and Phylogenetic Covariance Matrix P, as well as the relation between them.

The full phylogenetic likelihood can be written as

$$|2\pi(R \otimes P)|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T (R \otimes P)^{-1}(x-\mu)}$$

Which is the recognizable probability density function of a multivariate normal distribution. The $\otimes$ symbols represents the Kronecker product, which multiples R and P to form a covariance matrix. Meanwhile, x represents the states at the tips, stacked into a vector according to the order appropriate that in which the Kronecker product is taken. The symbol μ likewise represents the state at the root of the tree, repeated also in the appropriate order.

The two algorithms used here to diagonalize the matrices R and P are Cholesky factorization and the Felsenstein Pruning algorithm, respectively. Cholesky factorization provides the decomposition $R = LIL^T$, where L is the lower Cholesky factor and I is the identity. The Felsenstein pruning algorithm, meanwhile, decomposes the phylogenetic covariance matrix P into $C^{-1}QC^{-T}$, where Q is a matrix representing the branch lengths of each independent contrast, and C is the matrix that transforms the tip characters into independent contrasts. This decomposition is often represented as a post-order tree traversal, but can also be expressed and performed linear algebraically, operating purely on the phylogenetic covariance matrix.

Substituting these relations into the above expression, we obtain:

$$|2\pi(LIL^T \otimes C^{-1}QC^{-T})|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T (LIL^T \otimes C^{-1}QC^{-T})^{-1}(x-\mu)}$$

Using the properties of Kronecker products, we can rewrite this as:

$$((2\pi)^{nm}|LIL^T|^m |C^{-1}QC^{-T}|^n)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T ((L \otimes C^{-1})(I \otimes Q)(L^T \otimes C^{-T}))^{-1}(x-\mu)}$$

Which can be further reordered and simplified:

$$((2\pi)^{nm}|LIL^T|^m |C^{-1}QC^{-T}|^n)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T (L^T \otimes C^{-T})^{-1}(I \otimes Q)^{-1}(L \otimes C^{-1})^{-1}(x-\mu)}$$

$$((2\pi)^{nm}|LIL^T|^m |C^{-1}QC^{-T}|^n)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T (L^{-T} \otimes C^T)(I \otimes Q)^{-1}(L^{-1} \otimes C)(x-\mu)}$$

$$((2\pi)^{nm}|LIL^T|^m |C^{-1}QC^{-T}|^n)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T (L^{-1} \otimes C)^T (I \otimes Q)^{-1}(L^{-1} \otimes C)(x-\mu)}$$

$$((2\pi)^{nm}|LIL^T|^m |C^{-1}QC^{-T}|^n)^{-\frac{1}{2}} e^{-\frac{1}{2}((L^{-1} \otimes C)(x-\mu))^T (I \otimes Q)^{-1}(L^{-1} \otimes C)(x-\mu)}$$

$$((2\pi)^{nm}|LIL^T|^m|C^{-1}QC^{-T}|^n)^{-\frac{1}{2}}e^{-\frac{1}{2}((L^{-1}\otimes C)x-(L^{-1}\otimes C)\mu))^T(I\otimes Q)^{-1}((L^{-1}\otimes C)x-(L^{-1}\otimes C)\mu)}$$

And since

$$C\mu = 0$$

and

$$(L^{-1}\otimes C)\mu = 0$$

We can further simplify this into:

$$((2\pi)^{nm}|LIL^T|^m|C^{-1}QC^{-T}|^n)^{-\frac{1}{2}}e^{-\frac{1}{2}((L^{-1}\otimes C)x)^T(I\otimes Q)^{-1}((L^{-1}\otimes C)x)}$$

As I and Q are both diagonal matrices, we obtain in the place of our original function a multivariate normal distribution pdf whose covariance matrix is a diagonal matrix, with trait vector x appropriately transformed. This allows us to compute the original density as the product of univariate normal densities (or sum of log densities), so long as we take care to propagate the appropriate quantities to the determinant.