PMM versus PGLS, as I understand it

Common modeling approaches: There are two really common approaches to questions like this: calculate Pagel's λ on a trait, or correct for phylogenetic correlation in residuals in a typical analysis (PGLS) and analyse that correlation along the way (also called λ). These are **not** necessarily identical: one is evaluating the correlation structure of a trait and one is calculating the correlation structure of residuals. Neither one of these is what we want because they are not the model we think is at play (see below) and because both rely on one trait value per species, but I think I should understand them to get anywhere. So let's do the PGLS, which I am copying from the second edition of *Statistical Rethinking.*

$$y \sim MVN(\mu, S) \tag{1}$$
$$\mu_i = \alpha + \beta * x_i \tag{2}$$

μ is a usual linear model. y is reponse data (one per species), and S is a covariance matrix with as many rows and columns as species. In ordinary regression this takes the form:

$$S = \sigma^2 I \tag{3}$$

where I is just an identity matrix (all 1s) so we can ignore it. In PGLS we replace S with the phylogenetic covariance matrix (Σ). **You have to make sure of a few things**: in many packages the phylogeny must go in as correlation matrix (this makes the diagonals 1s and the off-diagonals the correlation across species due to evolutionary history) and make sure the rows and columns are in the same order as the species will be ordered numerically.

One issue with this model is that is that forces the correlation structure you give it—it does not adjust the correlation structure at all; some iterations of PGLS in certain R packages do this and it is quite an important addition to avoid Type I errors. I believe this just involves estimating a value to multiply the off-diagonals of matix by such (I don't know the notation to represent that λ is just on the off-diagonals).

$$S = \sigma^2 (\Sigma * \lambda) \tag{4}$$

So many models set λ to 1, but it's definitely best to ask the model to estimate it (rather than assume the phylogenetic correlation strutcure if how you want to structure the residuals).

We have some code from Nate Lemoine on how to code a PGLS that helped me see the structure (I suspect there are several ways mathematically to get to the same answer, but walking through this example helped me). Instead of Σ, we'll talk about the full variance-covariance matrix (V) and we'll introduce a couples matrices ($Lmat$) of all 1s on the off-diagonals and 0s on the diagonal, I (identity matrix) is the reverse, both are the same dimensions as V. This is bad notation (but it lets me follow the code), but $v\lambda$ and $v\sigma$ are single parameters:

$$v\lambda = (\lambda * Lmat + I). * V \tag{5}$$
$$v\sigma = \sigma^2 * v\lambda \tag{6}$$
$$y \sim MVN(\mu, v\sigma) \tag{7}$$
$$\mu_i = \alpha + \beta * x_i \tag{8}$$

So the first line: mulitples λ across the off-diagaonals, then adds the 1s to the diagonals, this yields:

$$\begin{bmatrix} 1 & \lambda & \lambda \\ \lambda & 1 & \lambda \\ \lambda & \lambda & 1 \end{bmatrix} \tag{9}$$

then multiples the whole thing element-by-element ($.*$ means element-wise matrix multiplication, aka Hadamard because everything in linear algebra has a name) by the VCV. The next line creates $v\sigma$ as that matrix multiplied by parameter σ^2. So we end up with:

$$\begin{bmatrix} \sigma^2 & \lambda\sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \lambda\sigma^2 & \sigma^2 \end{bmatrix} \tag{10}$$

And that is what goes in error part of $y \sim MVN(\mu, v\sigma)$.

PMM (phylogenetic mixed model): Okay, so what happens in PMM? Here's my understanding ...

$$y = \alpha + \beta x + a + e \tag{11}$$
$$a \sim normal(0, \sigma_P^2 \Sigma) \tag{12}$$
$$e \sim normal(0, \sigma_R^2 I) \tag{13}$$
$$\text{PGLS: } y \sim normal(\alpha + \beta x, \sigma_P^2 \Sigma) \tag{14}$$

... where α and β, respectively, are the intercept and the slope for the co-factor x, a is the phylogenetic random effect, and e is the residual error. Now, the two last terms are assumed to be normally distributed, with Σ as a phylogenetic correlation matrix , I stands for the relevant identity matrix. Our model, thus, assumes that phylogenetic effects are correlated according to the phylogenetic correlation matrix Σ. Note also that our model is estimating two variances: V_P is the variance of the phylogenetic effect and V_R is the residual error (environment effects, intraspecific variance, measurement error, etc.). [This text here is copied from Chapter 11: General Quantitative Genetic Methods for Comparative Biology, by Villemereuil & Nakagawa.]

A big difference often pointed out about PMM versus PGLS is that PGLS does not allow for non-phylogenetically structured error (but I think it sort of does once you scale the phylogenetic effect by λ, no?) and the PMM explicitly models other sources of error through e, and then in the PMM the strength of the phylogenetic effect is measured as:

$$\lambda = \frac{\sigma_P^2}{\sigma_P^2 + \sigma_R^2} \tag{15}$$

which is equivalent to just saying 'what proportion of the variance is due to phylogeny?' (This model has only a and e, while the Joly et al. has a, e and c.)

The multiple-values-per-species version of this (where you also have a value of x for each observation) involves within-group centering. The argument for this is to separate out the slope into its species-level and phylogenetic-level components. This makes the two slopes "perfectly orthogonal," but I hate it because it uses means you calculate outside of the model and if you have partial pooling then why are you trusting a mean you calculated without partial pooling?

One interesting thing I noticed about this model is that it estimates a **separate** intercept for the $y = \alpha + \beta x$ part of the model and I wonder if we need this? (Though the more I think on this, who is to say what is an intercept versus error in this model? They are both additive terms, one is just centered on zero I think.)

Will's PMM code

I have started work on some models that try to put the phylogeny on the slopes (which is what I want). Check out `ubermini.stan` and related R code. Here's the main breakthrough from today, which is just me learning math.

$$y = \alpha + \beta x + e \tag{16}$$
$$\beta \sim MVN(0, \Phi\Sigma) \tag{17}$$
$$e \sim normal(0, \sigma_y I) \tag{18}$$
$$\tag{19}$$

What does Φ look like? While here's β in the code currently:

```
b_force∼ multi_normal(rep_vector(0,n_sp),
diag_matrix(rep_vector(null_interceptsb, n_sp)) + lam_interceptsb*Vphy;
```

It says that my vector of slopes is multinormal centered around 0 (why zero? That's how Gaussian processes work, it somehow gets the 'centering' if you will from the $y \sim normal(\hat{y}, \sigma_y)$ bit of the model) and the variance should be the within-species variance on the diagonal, and the between-species variance on the off-diagonals.

Let α be `null_interceptsb` and λ be `lam_interceptsb`.

$$\Phi\Sigma =$$

$$\begin{bmatrix} \alpha + \lambda * Vphy & \cdots & \lambda * Vphy \\ & \ddots & \\ \lambda * Vphy & \cdots & \alpha + \lambda * Vphy \end{bmatrix} \tag{20}$$

When $Vphy$ is set to have 0s down the diagonal the above simplifies to this:

$$\Phi\Sigma =$$

$$\begin{bmatrix} \alpha & \cdots & \lambda * Vphy \\ & \ddots & \\ \lambda * Vphy & \cdots & \alpha \end{bmatrix} \tag{21}$$

When $Vphy$ is set to have 1s down the diagonal (e.g., `Vphy=vcv(phylo, corr=TRUE)`) the above simplifies to this:

$$\Phi\Sigma =$$

$$\begin{bmatrix} \alpha + \lambda & \cdots & \lambda * Vphy \\ & \ddots & \\ \lambda * Vphy & \cdots & \alpha + \lambda \end{bmatrix} \tag{22}$$

This additive part seems odd to me, I don't think it's what we want.

Some links:

This is an example of a PMM in Stan code. Varying slope with phylogenetic structure

Someone running a multiple measures model in BRMS with some queries. Not sure how useful this is.

Some random notes:

Cholesky decompostion: From what I understand, this is a trick when you cannot get the model to run with your covariance matrix, *Rethinking* says, it's way to "smuggle the covariance martix out of the prior."