Week 1

- Broad goal: Fix code.

    - ☑ Use less filtered dataset
    - ☐ Distance matrix argument to dist?
        - – Need both forms, so may or may not be useful
    - ☐ Check if phi is correct now, and see how is used
    - ☑ Figure out large alpha problem
        - – Seems the + and one loop beta fixed this?
    - ☐ Try a different covariate?
    - ☐ Figure out why geem code uses + instead of - on update
        - – I think + is correct. At least it converges.
    - ☐ Identify which covariate is causing the trouble.
    - ☐ Try removing the single OTU that seems to be contributing to problem.
    - ☐ Save everything to list correctly to look at later.
        - – Saving rho, omega, phi, beta, diff (abs), num iterations
        - – Also do residuals. Which residuals? Cross product ones? Squared ones? Standardized squared ones? Probably standardized but not squared. Should be distributed around zero. ?
    - ☑ Plot lists
        - – Plotted rhos, omegas, phis, diffs, not sure what else to plot.
    - ☑ Remove scalar update.
        - – Doesn't work. see journal
    - ☐ Write up algorithm for Yuan to go through? Ask to check.
    - ☐ Final run through before meeting of graphs and numbers
    - ☑ Move timing into function

Week 2

Main tasks:

- To-do from sheet
- Prepare presentation

☑ Wednesday email research update.

☑ Include more taxa: more zero-inflation

- Not great results. see journal omega = .99, rho = .1
- runtime of 20 min.

☑ Add computation time in model.

☑ Make presentation

☑ Figure out why adding the minus sign changes the code.

- Pretty sure this is because we are then adding or subtracting the λI incorrectly.

☑ Try changing the minus sign on larger dataset.

- Either immediate convergence or infinite alpha values on 1st iteration.....

☑ Set up automatic save outputs along with description

☑ Try with a medium sized dataset, like prevalence of .2?

☑ Save test datasets to files for easy loading

☑ Only include last beta output?

☐ Try different covariates (parasite burden/antibiotics/exposure)

☐ Think about genus level analysis/ phylogenetic tree for taxonomic units

☐ Choose α automatically (simple line search)

☐ Write up line-search steps and goals

☐ Understand how Hessian is calculated and expected values equal zero.

☐ Run on previous dataset from original paper.

☐ Rename file with algorithm

☐ Compare to independence GEE written already geeglm?

☐ Calculate sandwhich variance estimators for asymptotic significance tests

☐ See if new shared tree that Tom had is the same?

☐ Add residuals to output and look at plots

☐ Change initial value to something different?

- change initial rho to 10. Changes results (of at least ests of rho and omega)
- Try previous iterations rho and omega.
- See how beta estimates change
- Run code with initial value of 1 again to compare.

☐ Read numerical optimization book.

☐ Save convergence results of nls optim

☐ Make sure subtraction in ϕ is correct

Week 3

- General

    - ∗☐ Read numerical optimization book.
    - ∗☐ Understand how Hessian is calculated and expected values equal zero.
    - ∗☐ Run on previous dataset from original paper.
    - ∗☐ Rename file with algorithm
    - ☑ Take pictures of meeting notes.
    - ☐ Figure out how to send jobs to server

- Line search

    - ∗☐ Choose α automatically (simple line search)
    - ∗☐ Write up line-search steps and goals
    - ☐ Add condition if infinity or NA to halve gamma
    - ☐ Clean up code for speed and remove redundancy.
    - ☐ Look at notes from meeting about (different?) conditions
    - ☐ Code was set up incorrectly. Fix. Shouldnt be updating betas each time, just checking.
    - ☐ Save sum of absolute G for each iteration.

- Convergence

    - ∗☐ Save convergence results of nls optim
    - ☐ Change λ to be based on eigenvalues.
    - ☐ Save eigenvalues and look at ratio.
    - ☐ Test to see difference within iterations of if starting values (eg rho = 10) give different results.

- Correlation: Importance of ρ, ω

    - ∗☐ Compare to independence GEE written already geeglm?
    - ∗☐ Calculate sandwich variance estimators for asymptotic significance tests.
    - ☐ Keep ρ or ω the same throughout the algorithm

- Other

    - ∗☐ See if new shared tree that Tom had is the same?
    - ∗☐ Try different covariates (parasite burden/antibiotics/exposure)
    - ∗☐ Think about genus level analysis/ phylogenetic tree for taxonomic units
    - ☐ Add condition to make sure ρ, ω also converge
    - ☐ Email Tom with questions
    - ☐ Think about how hypothesis tests work
    - ☐ Read about GEE hypothesis tests

Week 4

☐ Writeup meeting notes

☐ Email Tom (response to my email) about more information

☐ Look into *resolving nodes*

☐ Select one Genus/Family to only have 10 ish ASVs. Filter dataset, and tree.

☐ Test against identity, fixed values of omega and rho.

Week 5

☑ Make gamma, lambda values arguments in the functions.

☑ Save final residuals in model.

☑ Convert initilization to one line list.

☑ See if weird jump still happens. figure this out.

- Seems like jump does not exist anymore.

☑ Add model call to result list

☑ Create diagnoistics plot function

- Patchwork outputs

☑ Clean up old code

☑ Save beta values again and look at changes.

☑ Run through with bug fixed. See what goes on. Write up results

☑ Figure out a way to compare beta values

- Using a line plot across iterations

☑ Add GEE values to plot.

☑ Check if gamma = 1 betas are same as gamma = .1 betas

- They are not.

☑ Figure out how to compare betas across different model runs (probably same line function?)

☐ Think about how to do simulations, since we need to have something to check against.

☐ Try to sum up phylogenetic tree and see if there are any problems.

☐ Email Tom about summing up the phylogenetic tree

☐ Look at residual matrix (cross residuals). See if close to zero.

☐ Use data that is just random common taxa.

☐ Try different covariates

☐ Consider moving functions to new files

☐ Convert all nls functions to matrices to make sure all entries match indeces.

☐ Update roxygen skeletons for functions

☐ Add final values of omega, rho and phi to patchwork plot

☐ Try independence

☐ Compare independence to geem beta values.

☐ Try fixed values of ω, ρ

☐ Figure out why flavo + 30p dataset stops in 2 iterations

- Figured out where the problem occurs, but not sure why it occurs...

☐ See how λ values change beta values.

☐ Explore convergence of optim loop.