

About the method

- **Q1:** About the skew symmetric matrix parametrization.
It is not clear to me how you use this to compute the derivative of the eigenvectors of C^{-1} (contained in matrix \mathbf{S}). What you told me is that if $\mathbf{S} = e^{\mathbf{X}}$ with \mathbf{X} skew-symmetric, then you can compute the derivative of \mathbf{S} w.r. to a change in individual elements of \mathbf{X} . Do you do this analytically? How?
The best would be that you write this in a detailed section in the SM. You can then correct the main text, in which I wrote equations using the Eulerian angles method.
- **Q2:** What happens if some eigenvalues of the $\mathbf{G}(\rho)$ matrices are degenerate? Then the gradient in equation (21) of the current MS diverges? How do you deal with this?
For instance, in the case of a perfectly balanced tree, this would happen all the time I suppose?
- **Q3:** Computing the eigenmodes $\{\lambda, \vec{u}\}$ of matrices $\mathbf{G}(\rho)$ for all eigenvalues ρ of the inverse correlation matrix is the computationally intensive part of this, right?
I suggested earlier that instead of recomputing those everytime, you update them with a Euler step using the gradient in equations 20 & 21 of the current MS. Every few steps, you could recompute them accurately. Have you tried this?
- **Q4:** Could you summarize the optimization technique you use in a few lines in the main text? You say it's a quasi-Newton method and give a reference. I know what a Newton method is, but am unsure what a quasi-Newton method is.

About the results

- **Q5:** Could you detail a bit more how you build the tree? Is it a fully balanced one?
- **Q6:** What is the typical runtime for the trees / sequences that you use? What I understand is that you use $2^{10} = 1024$ sequences of length $L = 10$. What if you use longer sequences?
- **Q7:** In figure 5, that is the PPV curves, why do you show only 16 predictions? With $L = 10$ I expected 45 predictions?

More general questions (for Martin too)

- **Q8:** I understand that right now we are using a fully balanced tree (even though it's not specifically written in the text, so I am not sure). Should we use a more "realistic" tree, e.g. something sampled from a coalescent model? It would be fairly easy to do.
An unbalanced tree could have cool effects. For instance if a part of the tree is very "bushy", having lots of closely related leaves, then our method could correct this. Whereas a naive estimation of couplings would be very biased.
- **Q9:** Should we vary the amplitude of the couplings? Or should we stick with the same strength of (equilibrium) correlations?
- **Q10:** Important question: what plots do we want?
Right now, we have the following scatter plots for 30 repetitions of the inference, and for three values of γ (with typical tree branch length staying the same):
 - Inferred γ vs real γ .
 - Slope of linear regression of inferred vs real couplings, vs Frobenius norm between inferred and real couplings.
 - Same as above, with Pearson correlation on the x -axis.
 - PPVs

I was suggesting in a previous round of questions that we could try to have something like "quality of inference" (Frob. norm, Pearson, ...) vs $\gamma\Delta t$. We could then see the regimes in which our thing is useful (intermediate strength of phylogeny I suppose).

Any other ideas are welcome.

- **Q11:** I realized that I dropped the case of balanced binary trees with equal branch lengths. I suggest that we have it as a supplement, since we are not using it for simulations? (It's a nice problem though, so we should keep it I believe).