Lecture #16 – Nonlinear Bayesian and Inverses

In this final lecture, we look at Bayesian and nonlinear inverses. Sadly, due to lack of time, we skipped the material on older, basic nonlinear methods in Chapters 9 and 10 in ABT, and went straight to Bayesian methods in Chapter 11. I would heartily recommend you also read those two chapters if you have the book and some free time!!

ABT state things nicely concerning the difference between the "classical approaches" and the Bayesian. "In the classical approach, there is a specific but unknown model … that we would like to discover. In the Bayesian approach, the model is a random variable and the solution is a PDF for the model parameters." Deterministic versus probabilistic – a big difference! But, that being said, the answers can look very much the same in some cases! So what's happening here? Let's see.

First, ABT show Bayes' Theorem (Theorem11.1) which is the foundation of the method. It is simply written as $q(m|d) = f(d|m)p(m)$. In this brief form, $p(m)$ is our a priori knowledge of the model, $f(d|m)$ is the "conditional probability" that we will get data d given a model m, and $q(m|d)$ is the "posterior distribution" (the answer) giving the probability of getting model m from data d. (We've ignored the normalization constant for now). The trick to this technique is forming the f(d/m) and p(m) and then doing a multidimensional integral. ABT show a nice example of how this works using an "uninformative prior", i.e., no prior knowledge of m, and a Gaussian $f(d_i|m)$. In this case, the Bayesian gives our old least squares solution!! This is because the "fitting error" to the data is incorporated in $f(d_i|m)$, see Eq. 11.10, and so we're really just looking at least squares being incorporated in a different equation form!

ABT give a nice example of how to use Bayesian methods via the simple example of repeated weighings of an object, including error! The use of priors and the power of having normal distributions are particularly emphasized.

In section 11.3 the idea that a normal prior distribution gives a normal posterior distribution is used to deal with multivariable cases (many model parameters). This is a rather simple matrix generalization of the single model parameter case. A clear example is again presented.

The final topic treated in ABT Section 11.4 is the use of maximum entropy methods to select a prior distribution subject to available information. The prior distribution is a key piece of the Bayesian inverse, and maximum entropy (i.e. maximum randomness) and other entropy metrics give one good, physically motivated way to construct a prior. Again, some rather simple and clear examples are given. A final "epilogue" in ABT also discusses when to use Bayesian methods, the answer being that they are at their most effective for ill conditioned problems, as they avoid the bias that regularization methods invariably introduce.

As an example of actual "acoustical oceanography" research using Bayesian methods, a paper by Stan Dosso and Charles Holland is provided. In this paper, a Bayesian approach to inverting bottom reflectivity data for seabed elastic medium properties (not just fluid!) is pursued. The theory section looks pretty much like the ABT Chapter 11 development, though with some added complexity, as the

inherently nonlinear problem tackled requires some numerical methods, including simulated annealing searches, which we will discuss soon! The answers one gets, shown in Fig. 5 of the paper, are of course PDF's. Interestingly, some are flat-liners,'' which simply means that the data have no sensitivity to those particular model parameters (as the priors were flat here.)

Our final topic in the course is nonlinear inverse methods, particularly simulated annealing (SA) and genetic algorithms. These are, in my view, full parameter space, iteration of forward model searches, but with a twist or two that let them avoid being fully brute force searches for the global minimum of a cost function. They are computer intensive.

For reading material on SA, I have excerpted an old chapter of "Numerical Recipes" and also cited a MATLAB Central webpage. The Numerical Recipes section has a nice description of an old, but basic, algorithm that SA uses to avoid "brute searches", the so called "Metropolis Algorithm." This algorithm is based on thermal physics (the Boltzman probability distribution) and the key feature of using this is that it allows some small amount of uphill searches, as well as the usual "going downhill fast" searches one usually does to find minima. By doing this, it allows an escape from trapping by local mimima, which is the occupational hazard in multidimensional searches. Exactly how one conducts this search is a bit of a "hands on" art form, as you have to fuss a bit with temperatures, configuration changes, numbers of iterations, and convergence criteria, but still – it's not too hard, and much canned software is now available. SA is a pretty popular technique! The Numerical Recipes chapter also has a very, very nice example in it – the famous "traveling salesman" problem. This can also be seen as a MATLAB demo, and you should try it – it is pretty cool to see the "trial paths" rearranging on the screen in real time! And, oops! Almost forgot that there was a WIKIPEDIA article on SA included as well. WIKI may not be "peer-reviewed," but if you realize that and maintain some healthy caution, you can actually find some very nice basic material there, like this write up!

The second, very popular, nonlinear inverse/search technique is the Genetic Algorithm (GA). It is a search technique based on evolutionary biology, as opposed to thermal physics. (Funny how we've copied natural processes here – annealing and evolution!). I have to admit that basing a search on "survival of the fittest" and mutations and gene mixing seems odd, but it works well and is widely used. Again, there are many, many software packages available for this technique (including MATLAB), and overall it is, like SA, not so hard to use. The WIKIPEDIA article gives a nice overview of GA and its many variants. After that, you have to get your hands dirty on a specific problem!

The final "parting shot" on nonlinear methods is a paper by Collins, Kuperman, and Schmidt (yes, Henrik Schmidt from MIT again!) on "Nonlinear inversion for ocean bottom properties." In this paper, the authors use the data that Lynch (me), Frisk (yup, same one) and Rajan collected in the Gulf of Mexico, and which has been discussed before. This paper is a very nice "basic/how to do" for SA inversions, as it carefully explores how you can use various data variants (the pressure field, modal wavenumbers, etc.) by choosing appropriate cost functions, and what the details of the inversion (schedule, convergence, etc.) look like. Their results are an improvement over our previous, linear inverse ones, in that SA inverse can easily include shear, the continuum, and of course, nonlinearity! This paper is 20 years old now, and today doing this formerly hard calculation is more like a term project – which points to some nice progress in the field.

2.682 Acoustical Oceanography

Spring 2012