I'm going to talk about a parameter identification problem for the so-called shallow shelf approximation from ice sheet modeling. The shallow shelf approximation is a model for ice flow applicable in certain areas of ice sheets such as ice streams and shelves. I originally had a slide with the SSA equations on it, but thought better of it, and I think that the following cartoon equations will suffice for discussion. We have two Laplacians coupled by low order terms and known forcing terms on the right-hand side. The variables u and v are components of ice velocity, while γ, which I'll call the bed strength, is the parameter of interest. This term is responsible for the shear stress at the base of the ice. If the denominator were not here, we would have a viscous bed model with stress proportional to velocity, and γ would be the constant of proportionality. As written, we can interpret γ as a yield stress: the bed can impart a shear stress up to the yield stress to keep the ice from sliding; where the ice is sliding the basal shear stress is at the yield stress.

At any rate, the forward model is: given a bed strength, determine velocities. The inverse problem goes the other way: given measurements of surface velocities, infer the bed strength.

The inverse problem is ill-posed for two reasons. The first is a counting problem: there is a scalar parameter that needs to account for a vector field of observations. So the inverse problem is overdetermined. This is a comparatively mild problem. The more insidious one is that it is discontinuous: errors in the observations can be magnified arbitrarily largely by the inverse map, and I'll say more about this shortly.

Now, forms of this inverse problem goes back nearly 20 years to the pioneering work of Doug MacAyeal who introduced the use of control theory methods. The idea is to minimize the misfit between observed and model velocities among all possible bed strengths, and the standard practice seems to be to use steepest descent as the optimization algorithm.

Given that this problem has been studied for so long, it is reasonable to ask 'what more is there to say?' I'm interested in doing parameter identification on full Greenland sized problems in collaboration with the Parallel Ice Sheet Model project, so I'm concerned about performance of the inverse method, and steepest descent leaves room for improvement. I'm also interested in a mystery: the mathematics literature on inverse problems pays a lot of attention to problem 2, while it is hardly ever mentioned in the ice literature. It is clear how introducing the functional J addresses problem 1, but why hasn't problem 2 an issue?

This slide is a demonstration of problem 2 manifests itself. This is a picture of a member of a family of exact solutions of the SSA due to Christian Schoof. These are flows through a cross-section, and hence are essentially 1-dimensional. On the top is the transverse ice velocity and on the bottom are the yield stresses. We're looking along the cross-section, and the units of the the variables are dimensionless. For x-coordinates larger than about 2 or so, the yield stresses are sufficiently large that the ice is genuinely frozen to the bed.

The smoothing nature of the forward map can be seen by adding some wiggles to the yield stresses. These modified yield stresses generate essentially the same surface velocities; the wiggles are damped. The higher the frequency of the wiggles, the greater the damping. This is the bane of solving inverse problems: when you go backwards, this damping becomes magnification, and it knows no bound. You cannot go about solving the inverse problem exactly. Instead, additional information must be used to select one solution over another. In this case, both yield stress fields are consistent with measurement, but I think we would reject the second on on the basis that the additional structure is not needed to explain the surface measurements.

This leads to the regularization method that I would be most excited about: I would like to find the least featured yield stresses that are consistent with observations. This can be phrased formally, for example, by minimizing some kind of norm of γ subject to the constraint that when you send in through the forward model, the resulting velocities are sufficiently close to the observations. The additional information is then encoded in the choice of norms X, Y, and the desired level of closeness δ. In selecting δ, one has to be aware that the sources of error are not just in the measurements, but also in the model. A variation on this strategy would be to specify not that the norm of γ be minimized, but rather its difference from some initial estimate γ_0, in which case we would be introducing one more piece of a-priori information. If any of you know how to do this well, I'd be delighted to talk with you.

There are some classical, more tractable methods of introducing a-priori information. The first is Tikhonov regularization. In it we perform an unconstrained minimization of the misfit with an appended penalty functional. Minimizing the mixed functional prevents finding a solution with low value of J that leads to an excessively high penalty term. The hard part about using this method is that the trade-off parameter ν needs to be selected. Although there is a literature of various schemes for automatically selecting one, they add computational complexity, and I find the comparative directness of the so-called iterative methods appealing. The additional information here comes from specifying an initial estimate γ_0 along with a desired level δ of misfit. These methods proceed by determining a sequence of search directions h_k. For each iteration, the misfit functional is minimized along the search direction, and we stop at the first iteration that is consistent with measurement. Clearly something has to be special about the search directions for this kind of approach to work. Miraculously, for a large number of methods, it is easy to select search directions that are initially less featured and become more featured as the computation progresses. Hence this method is a poor man's version of the constrained minimization algorithm I'd rather be solving.

I was going to try to explain how it is that these search directions can be found, but I'm low on time, so let me just say that the methods based using the gradient of the forward operator (including steepest descent and conjugate gradients) naturally have this property. There is a good explanation for it in terms of wiggles and self adjointness.

So we arrive at the last stop on our tour of regularization techniques, the iteratively regularized Gauss-Newton method. To understand it, let's start with the standard Gauss-Newton method for performing a nonlinear least-squares minimization. It is an iterative method, and at each iteration the nonlinear forward map is replaced with its linearization. This leads to a sequence of quadratic functionals that can be minimized exactly by solving a linear problem. The difficulty with using this kind of method with inverse problems is that full minimization of one of the quadratic functionals is just as ill-posed as the nonlinear problem. The nonlinear ill-posed problem has been reduced to a sequence of linear ill-posed problems. The standard approach to deal with this problem is know as the iteratively regularized Gauss-Newton method. In it, the intermediate functionals are minimized using Tikhonov regularization using a sequence of trade-off parameters ν_k.

Finally we come to the method that we are using that I'll call incomplete Gauss-Newton. I'm not aware of it being in the literature, but it seems so natural I suspect it's already known to somebody – if it's one of you, please talk to me. Anyway, it is based on the Gauss-Newton method as above, but it deals with the intermediate ill-posed problems using the conjugate gradient method and incomplete minimization. We start with an initial estimate γ_0 and a desired level of misfit. At each iteration, the linear ill-posed problem is regularized by removing only a fraction θ of the remaining misfit. Once a search direction has been found, the rest of the algorithm proceeds in the natural way. We also perform some management of the θ using trust region ideas: if the observed decrease of J is not close to the one predicted by the quadratic models, we shrink θ.

One of the features of this method that I find most appealing is that the a-priori information you specify is very natural: an initial estimate and a desired level of misfit. It turns out that the method is relatively speedy as well. Compared to the nonlinear conjugate gradient method, many of the iterations are shunted from solving nonlinear problems to solving linear ones. A bonus artifact of the way the intermediate linear ill-posed problems are solved, our linesearches often meet the Wolfe-Powell conditions on the initial guess and frequently only one nonlinear function is solved in an iteration.

This is a graph put together by a graduate student, Marijke Haberman, who has been working with IGN and the SSA and has been investigating the importance of taking into account the stopping criterion when using it. It shows the relative performance of steepest descent, nonlinear conjugate gradients, and incomplete Gauss-Newton in solving an particular inversion for the SSA. The x-axis is the outer most iteration number. For each algorithm, the outermost iteration is dominated by the line search where nonlinear problems are solved, so this is a good proxy for speed. The y-axis is the level of misfit, and the cyan line is the desired level of misfit (and hence the stopping condition). I think this graph explains why folks using steepest descent didn't run into trouble with wiggles sooner – the standard stopping criterion is to wait until no much change has been detected in the missfit, and steepest descent just doesn't have the oomph to get you into the region where you might get trouble from overfitting error. On the other hand, it would be easy to underfit with steepest descent, and there seems to be some evidence of this.

This is an example of inversion for basal shear stresses of a synthetic ice stream using steepest descent. On the left are the true basal stresses and the greenish ones are the differences between computed and actual stresses starting from various initial estimates. The data is very clean, with some minor interpolation errors having been introduced by the authors by performing their inversions and forward computations on different grids. I believe that the interpretation of this has been that there are multiple minima, and you arrive at a different minimum depending on your starting point. Interestingly, the terminal misfits vary from about 3m/a to nearly 9m/a, and I think what's happened here is that steepest descent gave up too soon.

This is my approximation of this same experiment, using the geometry data generously provided by Ian Joughin. The misfit in all cases has been reduced to 1m/a, and we do not see substantial differences between the various computed basal stresses. This is promising, and there is hope that by using a more powerful inverse method, there is the possibility of obtaining yield stresses that are as detailed as measurements allow. That said, we are only now starting to put the method to work on real data, and it remains to be seen what improvements, if any, there are in the detail of the reconstructions using IGN. Nevertheless, the additional speed of the method also leads to the possibility of using it for larger scale problems (as in the aformentioned PISM work) as well as for higher order models.

In the interest of full disclosure, it seems fair to point out some limitations of IGN. The first that comes to mind is its lack of flexibility with regards to the norms involved. For various technical reasons, they all need to be Hilbert spaces (such as L^2). This is in contrast to Tikhonov regularization where you are free to add, for example, a L^1 penalty norm which is more forgiving of discontinuities. This is not a fancy method that simultaneously performs a pseudo-optimization while solving the nonlinear problem. Nevertheless, it has a high utility to complexity ratio: folks using steepest descent have essentially all the tools needed to switch to IGN. I mention that there are no proofs that this works, which is in contrast to the large mathematics literature which proves, for example, that iteratively regularized Gauss-Newton is a regularization strategy.

I had hoped to talk about another experiment that bodes well for the use of the method in, for example, Greenland, where there are small regions of interesting flow surrounded by seas of slower ice, but it seems I'm out of time, and I'd be happy to talk one-on-one later to anyone who is interested.

Thanks.