Extended Abstract ECMI 98

Solving parameter estimation problems using regularized nonlinear least squares with applications

Mårten Gulliksson

Umeå University, 091 87 Umeå, Sweden
Tel +46 90 166548, Fax +46 90 166126, email marten@cs.umu.se

Keywords: Regularization, least squares, chemical processes, optimization, L-curve, ill-posed, conductivity

Introduction

The nonlinear least squares problem may be used to solve a large class of problems ranging from nonlinear system of equations to parameter estimation and surface fitting. In recent years much interest has been focused on problems where the solution is either not unique or may be very sensitive to perturbations in input data.

One class of ill-posed problems is inverse problems appearing in many different engineering applications. An inverse problem consists of a direct problem and some unknown function(s) or parameter(s). In many cases the solution does not depend continuously on the unknown quantities. A typical ill-posed problem is when the task is to determine these unkowns given measured, inexact, data.

Given such an ill-posed problem it is a good idea to reformulate the original problem into a well-posed problem that gives a solution that is neither too large nor giving a large residual. For our applications we have chosen to investigate the use of Tikhonov regularization and Gauss-Newton methods and we will further disuss this approach.

Tikhonov Regularization and Gauss-Newton Methods

Consider the nonlinear least squares problem

$$\min_x \ \frac{1}{2}\|f(x)\|_2^2 = F(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^m$ is at least twice continuously differentiable and $\|\cdot\|_2$ is the 2-norm.

The Tikhonov regularization for the nonlinear least squares problem consists of solving the problem

$$\min_x \frac{1}{2}\|f(x)\|_2^2 + \frac{1}{2}\lambda\|x - x_c\|_2^2, \tag{2}$$

where $\lambda > 0$ is the regularization parameter and x_c is some center ideally chosen as the critical point of interest but often just as zero. Choosing λ large enough we can always get a well posed problem since the Hessian is positive definite. This makes Tikhonov Regularization applicable regardless of the type of ill-posedness. However, the actual implementation should take into account if the problem is exactly rank deficient at the wanted critical point. The difficulty is to choose λ as small as possible and at the same time getting the solution and the residual of reasonable size.

The general idea is to apply a Gauss-Newton method on (2) with a suitably chosen sequence of regularization parameters $\{\lambda_k\}$. This is in many cases a delicate business since there is always the trade off between efficiency, size of the approximation $\|x_k\|$, and size of the residual $\|f(x_k)\|$. The approach taken here is to use the L-curve to determine the regularization parameter. In the case the Gauss-Newton method has slow convergence a specially designed Quasi-Newton method is used also on the Tikhonov problem.

The L-Curve

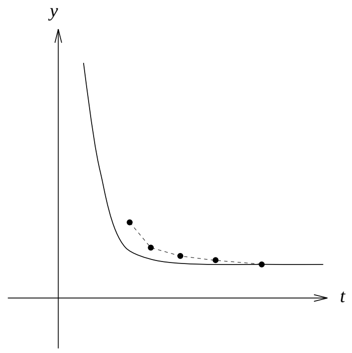
We have the following definition of the L-curve.

Definition 0.1. *Let $x(\lambda)$ solve problem (2), i.e.,*

$$x(\lambda) = \arg\left\{\min_x \; t(x) + \lambda y(x)\right\}, \quad \lambda \geq 0,$$

where $t(x) = \|f(x)\|_2^2/2$ and $y(x) = \|x - x_c\|_2^2/2$. The L-curve is the curve $(t(x(\lambda)), y(x(\lambda)))$.

A typical L-curve for the special case of nonlinear least squares is shown in Figure 1.



Figur 1: L-curve.

The corner of the L-curve may be used to find a reasonable regularization parameter. In the nonlinear case there may be several such corners but they are all found by minimizing $t(x(\lambda))y(x(\lambda))$. This fact together with the property that any approximate solution lies above the L-curve makes the L-curve very useful.

Determination of Conductivity

A common problem in applications is the determination of the conductivity $a(x)$ in the differential equation

$$\begin{aligned} -(au_x)_x &= f \\ u(0) = u_0, &\quad u(1) = u_1 \end{aligned} \tag{3}$$

from measurements of u in $]0, 1[$. This is an ill-posed problem since

$$a(x) = \frac{1}{u_x(x)}\left[a(0)u_x(0) - \int_0^x f(s)\, ds\right]$$

and attaining $a(x)$ requires the differentiation of u which is an ill-posed problem in the presence of inexact data of u.

The implementation of a Gauss-Newton method requires the calculation of $u = F(a)$ and the Jacobian, i.e., the differential equation (3) has to be solved numerically. This is performed with a linear spline ansatz $\sum_{j=0}^{n+1} u_j \phi_j(x)$ for u and the same type of ansatz $\sum_{j=0}^{m+1} p_j \phi_j(x)$ for a.

In an implmentation of the Gauss-Newton method for this problem one may either have a finite dimensional approach or an infinite dimensional formulation of the method. In the finite dimensional case one explicitly calculates the Jacobian in discrete space and use "standard" optimization techniques on the regularized problem. The model algorithm looks something like this.

0. Discretisize the problem.
   Find initial values of the solution, the regularization parameter, and the center.

1. Do a convergence test.
2. Determine the Jacobian (and second derivatives).
2. Iterate with the Gauss-Newton method on the regularized problem.
3. Update the L-curve (and a-curve) and choose the regularization parameter.
4. Go to 1.

The use of the L-curve (and a-curve) may be more or less sophisticated and automatized but for effeciency we have chosen to use a convex spline approximation.

Another approach is to keep the infinite dimensional problem intact as long as possible and use a variational formulation of the form

$$\sum_{j=1}^{m+1} [\alpha_k < \phi_i, \phi_j >_{H_1} + < F'\phi_i, F'\phi_j >_{L_2}] p_j \tag{4}$$
$$= - < F - y, F'\phi_i >_{L_2} - < a_k - a_0, \phi_i >_{H_1} .$$

The Frechet derivative F' at a_k is attained by solving another differential equation.

We will report on computational experiments for both the finite and infinite approach.

Determination of Moisture Content in Paper

A model for the moister content in a flat arc of paper is the system of partial differential equations

$$\frac{\partial w_p}{\partial t} = \frac{\partial}{\partial x}(D_p \frac{\partial w_p}{\partial x}) - N_p(kw_p - w_f)$$
$$\frac{\partial w_f}{\partial t} = D\frac{\partial}{\partial x}(D_f \frac{\partial w_p}{\partial x}) + N_f(kw_p - w_f)$$

where w_p, w_f is the water content in the pores and fibres respectively of the paper. It is of interest to determine the conductivity parameters D_f, D_p from measurements in order to find a suitable model. This is a typical inverse problem generally ill-posed in the same way as the conductivity problem stated earlier. However, the approach using a regularized problem with Gauss-Newton is still applicable and only the actual discretization differs.

Modelling a Continuous Digester

A really difficult problem is to determine a good model of a continuous digester. Most models are of the form

$$\frac{\partial u}{\partial t} + D\frac{\partial u}{\partial z} = f(u)$$

where z is the spatial variable along the digester, u is a vector of state variables such as temperature and concentrations, and $f(u)$ is a nonlinear vector function.

We are currently investigating the use of Collocation and Gauss-Newton methods to determine different model parameters of the digester from measurements. The parameter estimation problem is most probably ill-posed but this is yet to be seen.

Artificial Neural Networks

Our final example is the determination of weights in a feedforward neural network. The data is both simulated and from standard test examples. This problem is a highly nonlinear and overparameterisized least squares problem with many local minima.

Since this is, in general, a very large problem there is a need for iterative methods when using the Gauss-Newton method.