Extended Abstract ECMI 98

Regularizing nonlinear least squares with applications to parameter estimation

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, parameter estimation, optimization, L-curve, ill-posed

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.

In the optimization community these problems have often been treated as special cases not suitable for standard software. One reason is that regularity conditions at a local solution of the minimization problem are almost always imposed in order to get the methods to behave well close to a local solution. Stabilization is used in the methods in order to get sufficiently close to the well behaved local solution. Another reason is that the optimization problem is almost always formulated as a finite dimensional one ignoring the possible fact that the background (infinite dimensional) problem is ill-posed. Consequently, the terms ill-posed and regularization, refering to problems that are not well defined at a local solution, are seldom used by optimizers. However, ill-posed problems do indeed exist in optimization.

We will define what we mean by an ill-posed nonlinear least squares problem. This is simple enough using second order derivatives but for our problem there is more to say. The main reason is the structure of the Jacobian and the Hessian. By using this structure we can do a more informative analysis.

Another reason for looking more closely on the structure of the problem is that the natural method for solving nonlinear least squares methods is the Gauss-Newton method. In the Gauss-Newton method a linear least squares problem is solved in each iteration using the Jacobian. Obviously, there will be trouble using this method if the Jacobian is very ill-conditioned at the solution of interest. To handle this case we use Tikhonov regularization together with the L-curve.

The ill-posed unconstrained problem

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 first order KKT-condition for (1) is

$$\nabla_x F = J^T f = 0, \tag{2}$$

where $J = \partial f / \partial x$ is the Jacobian of f. A solution $\hat{x}$ to (2) will be called a critical point. For clarity we sometimes denote functions or derivatives evaluated at $\hat{x}$ with a hat, e.g. $\hat{J} = J(\hat{x})$.

Obviously, if the Hessian $\nabla_{xx}^2 F = J^T J + \sum_{i=1}^m f_i f_i''$ does not have full rank or is very ill-conditioned at the critical point of interest we have an ill-posed problem. Suprisingly enough this will *always* be the case when when J is of rank $r < n$ in a neighbourhood of the critical point $\hat{x}$. To be more specific we state the following theorem.

Theorem 0.1. *Let $\hat{x}$ be a critical point and the rank of J equal to $r < n$ in a neighbourhood of $\hat{x}$. Then $\nabla^2_{xx}F(\hat{x})$ is a matrix of rank $r < n$ with its nullspace containing the nullspace of $J(\hat{x})$.*

Examples of such ill-posed problems are underdetermined nonlinear system of equations and nonlinear total least squares. We may conclude that having J rank-deficient makes (1) an ill-posed problem in the sense that (2) does not have a unique solution (but a local minimum to (1) may exist though). For the constrained nonlinear least squares problem with rank deficient constraints we have shown results corresponding to Theorem 0.1.

The strong dependence between the Jacobian and the ill-posedness of the problem is partly inherited for the case the Jacobian is very ill-conditioned in a neighbourhood of the critical point. Consider the SVD of $J = U\Sigma V^T$ and assume that we have a gap in the singular values such that $\sigma_{r+1}/\sigma_r = \epsilon \ll 1$. We will also, for simplicity, assume that we have used the SVD to transform our problem such that we get a new $J = \Sigma$. We partition $J = \Sigma = \text{diag}(\Sigma_1, \Sigma_2)$ with Σ_1 containing the first r singular values and define $J_1 = \text{diag}(\Sigma_1, 0)$ and $J_2 = \text{diag}(0, \Sigma_2)$. Then $J = J_1 + J_2$ with J_1 of rank r and $J^T J = \text{diag}(D_1^2, D_2^2)$ where $D_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$ and $D_2 = \text{diag}(\sigma_{r+1}, \dots, \sigma_n)$.

Consider the Hessian

$$\nabla^2_{xx}F = J^T J + \sum f_i f_i'' = (J_1^T J_1 + \sum_{i=1}^{r} f_i f_i'') + (J_2^T J_2 + \sum_{i=r+1}^{m} f_i f_i'').$$

Define $\sum_{i=1}^m G_i^{(1)}$ as the second derivatives corresponding to J_1. Note that this is generally not any part of the second derivatives $\sum f_i f_i''$. It can be shown that $f_i f_i'' = G_i^{(1)} + \mathcal{O}(\epsilon)$ making the first term an ill-conditioned matrix since $J_1^T J_1 + \sum_{i=1}^r G_i^{(1)}$ is exactly rank deficient according to Theorem 0.1 above. The second term is small only if $\sum_{i=r+1}^m f_i f_i''$ is small giving a simple condition when then Jacobian does not contain information enough to determine if the problem is ill-posed. An important case is when $f = f_1 + f_2$, the Jacobian of f_1, J_1, has rank $r < n$ and f_2 is some noice (function) smaller than the relative gap ϵ.

Tikhonov regularization

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

$$\min_x \frac{1}{2}\|f(x)\|_2^2 + \frac{1}{2}\lambda\|x - x_c\|_2^2, \tag{3}$$

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 $\nabla^2_{xx}F = J^T J + \lambda I + \sum_{i=1}^m f_i f_i''$ 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 approach taken here is to use a Gauss-Newton method for the Tikhonov problem and using 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.

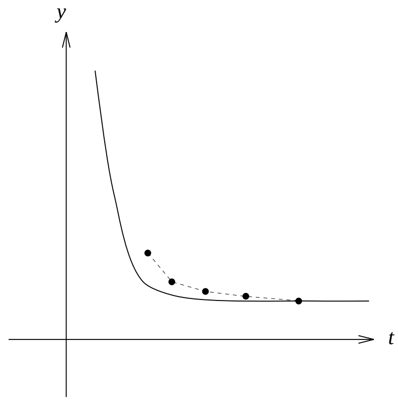
The L-curve

We make the following definition of the L-curve.

Definition 0.2. *Let $x(\lambda)$ solve problem (3), 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 defined as the curve $(t(x(\lambda)), y(x(\lambda)))$.



Figur 1: L-curve.

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

We will show that the L-curve $(t(x(\lambda)), y(x(\lambda))$ defines a strictly decreasing convex function $y(t)$ with the derivative $dy/dt = -1/\lambda$. Another very important property of the L-curve is that for any $\tilde{x} \in \mathbb{R}^n$ the corresponding point on the L-curve $(t(\tilde{x}), y(\tilde{x}))$ *always lies above the L-curve* as indicated in Figure 1. This fact will be proved using two related minimization problems that further motivates the use of the 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 in the nonlinear case.

Problems in parameter estimation

We will show results for several finite and infinite dimensional parameter estimation problemsa using our Gauss-Newton method on the Tikhonov problem. The regularization parameter has been chosen with the help of the L-curve.

Our first example it the Urysohn integral

$$\int_0^1 k(s, x(t))\, dt = y(s), \quad 0 \leq s \leq 1$$

where the kernel $k(s, x(t))$ is nonlinear in $x(t)$. The function $x(t)$ is a vertical profile of a physical quantity that can not be measured directly. The kernel function expresses the transmissibility properties of the layer under consideration with respect to the rays or waves passing through a layer and yielding the measurements $y(s)$ depending on the angle of incidence or wave length s.

We have chosen different solutions $x(t)$ giving an exact $y(s)$ The problem is discretized by approximating x in a spline basis and using the trapezoidal rule for the integral. Finally, we add noice to get a zero residual nonlinear least squares problem.

Our second 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 finite dimensional nonlinear overparameterisized least squares problem with many local minima.

Finally, we will try to determine the conductivity or diffusion coefficient $a(x) > 0$ in

$$-\frac{d}{dx}(a(x)\frac{du}{dx}) = f(x), \quad 0 < x < 1,$$
$$u(0) = u_0, \quad u(1) = u_1.$$

This problem is ill-posed since a depends on u_x. An exact solution u is chosen. The problem, i.e. a and u, is then discretized in the proper function spaces using finite elements. A finite dimensional nonlinear least squares problem is attained by adding noice to the discretized u.