Regularization & Preconditioning

Mauricio D Sacchi[1]

THEORY

Minimization of a quadratic cost function

Consider finding the solution of the following inverse problem

$$L\,m + \eta = d\,, \tag{1}$$

where m is a $M \times 1$ vector of model parameters (the image) and d indicates an $N \times 1$ vector of observations (data) contaminated by noise η ($m \in R^M, d\,\text{and}\,\eta \in R^N$). The matrix L has dimensions $N \times M$ and in our case represents a forward modelling operator. We will first consider matrices but bear in mind that we will replace L by any general linear operator. We attempt to solve our inverse problem by minimizing the following cost function

$$\begin{aligned} J &= J_d + \mu\, J_m \\ &= \frac{1}{2}||L\,m - d||^2 + \frac{1}{2}\mu\,||W_m\,m||_2^2\,. \end{aligned} \tag{2}$$

The first term of the cost function is the misfit J_d, a measure that represents fitting fidelity. The second term is the regularization term. The later is included to guarantee the stability and uniqueness of the solution. The regularization term is often chosen to impose desirable features onto the solution. The matrix (or linear operator) W_m represents weights applied to the vector of model parameters. At this point we introduce the following new variable

$$v \;=\; W_m\,m \tag{3}$$

we will also assume that W_m is invertible

$$m = W_m^{-1}\,v = P\,v\,. \tag{4}$$

Equation 2 can now be expressed as follows

$$\begin{aligned} J &= J_d + \mu J_m \\ &= \frac{1}{2}||(L\,P\,v - d)||_2^2 + \frac{\mu}{2}||v||_2^2\,. \end{aligned} \tag{5}$$

The following interesting points about equation 5 are in order:

- First, we have reduced the cost function 2 to the the so called standard form. The standard form corresponds to the cost function of the classical damped least-squares problems. One can turn a solver for 5 into one to solve 2. An algorithm such as `CGLS` can be used to find the solution v that minimizes 5 and then v can be used to obtain m.

[1]Email: `msacchi@ualberta.ca`
Signal Analysis and Imaging Group, Department of Physics, University of Alberta

- The solution that minimizes J is obtained by setting $\nabla_m J = 0$ which leads to

$$v = (P^T L^T L P + \mu I)^{-1} P^T L^T d \quad (6)$$
$$m = P\, v\,. \quad (7)$$

- In general, W_m is a high-pass operator (roughening operator). Therefore, P must behave like a low-pass operator (smoothing operator).

- The real advantage of using this formulation arises when we minimize J via iterative or semi-iterative methods (for instance, steepest descent or `CGLS`). In this case, the solution to the problem is attained by a finite number of steps (iterations) where the application of $L\,P$ or $(L\,P)^T$ to vectors in R^M and R^N dominates the cost of each iteration. It is essential to realize that in many imaging and signal processing problems we do not have explicit access to L. In other words, L is not a matrix; we do only have access to the action of the linear operators $L\,P$ and $(L\,P)^T$ on vectors $x \in R^M$ and $y \in R^N$, respectively.

Close form solutions of (2) and (7) are equivalent

The solution of (2) is (r: regularization)

$$m_r = (L^T L + \mu W_m^T W_m)^{-1} \; L^T d \quad (8)$$
$$(9)$$

The solution of (7) (p: preconditioning) is

$$m_p = P(P^T L^T L P + \mu I)^{-1} P^T L^T d \quad (10)$$
$$(11)$$

$m_r = m_p$ if P is invertible $P^{-1} = W_m$.

Quadratic misfit with non-quadratic regularization

Let us assume a non-quadratic regularization term such as the Cauchy criterion, the l_1 norm, or the Huber norm. These are often used to impose sparsity on model parameters.

l_1 regularization

We start by assuming the following cost function

$$J = J_d + \mu\, J_m$$
$$= \frac{1}{2}\|(L\,m - d)\|^2 + \mu\,\mathcal{R}(m) \quad (12)$$

where now R is the non-quadratic norm

$$\mathcal{R}(x) = \sum_{k=1}^{M} |x_i|$$

Evaluating $\nabla_m J = 0$ leads to

$$\begin{aligned} \nabla_m J &= \nabla_m J_d + \mu \nabla_m J_m \\ &= (L^T L\, m - L^T d) + \mu\, Q(m)\, m = 0 \end{aligned} \tag{13}$$

where $Q(m)$ is a $M \times M$ diagonal matrix that depends on model parameters

$$[Q(m)]_{ij} = \frac{\delta_{ij}}{|m_i|} \tag{14}$$

with

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j\,. \end{cases}$$

To void dividing by zero, we change the diagonal matrix Q as follows

$$[Q(m)]_{ij} = \frac{\delta_{ij}}{|m_i| + \epsilon} \tag{15}$$

We now have to solve a non-linear problem; the solution m depends on $Q(m)$. We can use the following iterative scheme, starting with $m^0 = 0$, we solve

$$(L^T L + \mu Q(m^{k-1}))m^k = L^T\, d\,. \tag{16}$$

Equation 16 is also the minimum of the following quadratic cost function:

$$\begin{aligned} J^k &= \|L\, m^k - d\|_2^2 + \mu\, \|W(m^{k-1})\, m^k\|_2^2 \\ W(m)_{ij} &= \frac{\delta_{ij}}{\sqrt{|m_i| + \epsilon}} \end{aligned} \tag{17}$$

We can now use the pre-conditioning trick to transform (14) into the standard form for the unknown v

$$[P]_{ij} = \delta_{ij}\sqrt{|m_i^{k-1}| + \epsilon} \tag{18}$$

$$J^k = ||\, P\, v^k - d||_2^2 + \mu\, ||v^k||_2^2 \tag{19}$$

$$m^k = P^{k-1} v^k \tag{20}$$

The algorithm can be used to retrieve *sparse* solutions. We often called this algorithm *Iterative Re-weighted Least-Squares* (IRLS). In our case, we have implemented IRLS via pre-conditioning. In Fourier-based interpolation/reconstruction equation (19) is can be iteratively solved via CGLS. From the iterative algorithm given in equations (18)-(20) it is clear that in each iteration you need to solve equation (19) and this is done ”on the flight” with CGLS. These tricks have been in used for solving Radon transforms and seismic reconstruction problems (Sacchi and Ulrych, 1995; Sacchi et al., 1998; Trad et al., 2003; Liu and Sacchi, 2004). Least-squares migration algorithms often use this style of preconditioning (Wang et al., 2005).

REFERENCES

Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, 1560–1568.

Sacchi, M. D., and T. J. Ulrych, 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, 1169–1177.

Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Transactions on Signal Processing, **46**, 31–38.

Trad, D., T. Ulrych, and M. D. Sacchi, 2003, Latest views of the sparse Radon transform: Geophysics, **68**, 386–399.

Wang, J., H. Kuehl, and M. D. Sacchi, 2005, High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the Western Canadian Sedimentary Basin: Geophysics, **70**, S91–S99.